DEPARTEMEN STATISTIKA, INSTITUT PERTANIAN BOGOR

Sidik Peubah Ganda

Dengan menggunakan SAS

Ahmad Ansori Mattjik & I Made Sumertajaya

2011

KAMPUS IPB DARMAGA, JL. MERANTI W22 LV 3

KATA PENGANTAR

Alhamdulillah, puji sukur kehadiratNya, karena atas ridha dan kehendakNya buku ini dapat kami tulis. Setelah mengajar Sidik Peubah Ganda sejak tahun 1985 sampai sekarang di Departemen Statistika, Fakultas Matematika IPB, dan atas dorongan dari kawan sejawat, kami berkeinginan untuk mengembangkan catatan kuliah menjadi tulisan yang dapat diterbitkan dalam bentuk sebuah buku. Pengalaman selama ini memberikan inspirasi macam dan bentuk tulisan yang diperkirakan dapat memberikan petunjuk dan tuntunan untuk mengertikan peubahganda pada mahasisiwa. Buku ini tersusun atas bantuan banyak pihak, untuk itu kami sangat berterimakasih kepada rekan sejawat dan mahasiswa bimbingan yang telah memberikan data hasl penelitiannya untuk digunakan sebagai contoh, memberikan kritik dan saran penulisan, serta dukungan moril maupun materi.

Kami sangat berhutang budi pada guru-guru kami yang telah memberikan jalan dan bimbingannya dengan sepenuh hati, tetapi karena keterbatasan kami mohon maaf tidak dapat menuliskan namanya satu persatu. Semoga kebaikan guru-guru kami tercatat senagai amal baik dan mendapat balasan dengan berlipat ganda kebaikan dari Allah SWT. Amien.

Kami yakin, masih banyak kritik dan saran yang diperlukan dari pembaca agar buku ini menjadi lebih bermanfaat. Untuk itu, kami mengucapkan terimakasih yang berlipat ganda. Meskipun demikian kesalahan dan kehilafan yang ada dalam buku ini tetap menjadi tanggung jawab kami.

Bogor, Juni 2011

Penulis,

Ahmad Ansori Mattjik & I Made Sumertajaya

DAFTAR ISI

1

1.Pendahuluan

Sesuai dengan perkembangan pengetahuan yang tidak ada kesudahannya, demikian juga dengan perkembangan dari metoda atau tatacara untuk interpretasi data. Bila semula hanya cukup dengan memperhatikan satu peubah atau karakter saja dari satu individu, sekarang berkembang dengan memperhatikan dua bahkan lebih banyak lagi peubah yang merupakan karakter dari individu yang sama. Seperti pengamatan pada padi, yang semula hanya memperhatikan produktivitasnya saja, kemudian berkembang dengan memperhatikan umur dan tinggi tanaman, sehingga diperoleh tanaman yang lebih pendek umur dan tingginya namun menghasilkan produktivitas yang tinggi. Berkembang lagi dengan memperhatikan rasa, warna, dan tempat tumbuh. Sehingga diperoleh padi yang sesuai di semua tempat atau hanya sesuai ditempat tertentu saja. Secara umum, karena merperhatikan banyak karakter dari individu yang sama, akan ada saling ketergantungan atau inter dependensi antar karakter yang diamati, atau dikenal dengan saling berkorelasi. Keadaan itulah yag menjadi fitur yang dapat membedakan antara peubah ganda dan peubah tunggal.

Buku ini ditulis terutama untuk menyediakan tehnik pengolahan data peubah ganda bagi para ilmuwan hayati yang pengumpulan datanya diperoleh dari suatu perancangan percobaan. Meskipun dapat juga digunakan untuk para ilmuwan dari bidang lain dengan penyesesuaian seperlunya. Sebagai pengantar ke analisis peubah ganda, materi yang

dicakup akan meliputi manova, analisi profil, analisis komponen-utama, analisis faktor, analisis gerombol, analisis diskriminan, analisis biplot, analisis korespondensi, analisis kanonik, analisis regresi peubah ganda, model persamaan berstruktur, penskalaan berdimensi ganda, dan analisis konjoin, disamping review aljabarmatriks.

Buku ini disusun dalam 17 bab dan disusun sebagai buku pegangan untuk mata-ajaran sidik peubah ganda bagi mahasiswa S1 pada tahun ke tiga atau mahasiswa S2 tahun pertama mayor statistika di Departemen Statistika, Fakultas Matematika dan IPA, IPB. Bagi mahasiswa S1, sampai dengan Bab 3 wajib dimengerti dengan baik sebelum mempelajari Bab lainnya. Bab 12 sampai Bab 17 adalah topik yang wajib diketahui bagi mahasiswa S2, sedangkan bagi mahasiswa S1 baik untuk diketahui akan tetapi diberikan sebagai pengetahuan tambahan saja. Dalam mempelajari buku ini, meskipun akan ada pengulangan untuk beberapa materi peubah tunggal seperlunya, tetapi untuk memperoleh hasil yang lebih baik, pembaca diharapkan telah mendapat pengetahuan satu semester untuk mata-ajaran metoda statistika, sampling, dan kalkulus, terutama tentang turunan sebagian dan integral, yang diperlukan untuk menghitung maksimum dan nilai harapan. Akan lebih baik lagi bila telah mempelajari aljabar matriks.

Kami berterimakasih pada Saudara Alfian dan Adhi, mahasiswa S3, yang telah membantu kami mengedit buku ini dengan seksama.

2

2.Review Aljabar Matriks (*Matrix Algebra Review*)

2.1. Pendahuluan

Pada beberapa bidang ilmu dan bidang terapan, seperti sosiologi, politik, ekonomi dan keputusan-keputusan bisnis, sebagian besar informasi diperoleh dari hasil analisis data. Seringkali data yang tersedia sangat besar, sehingga jika langsung diintrepetasikan akan cukup menyulitkan, kecuali dilakukan dulu tahapan peringkasan dan interpretasi yang sesuai. Seringkali juga, analisis yang tepat tidak dapat hanya mengandalkan perhitungan sederhana seperti rata-rata saja. Semakin kompleks struktur datanya, kebutuhan data analisis yang lebih konperhansip juga diperlukan.

Kerumitan dari sebuah gugus data muncul karena beberapa alasan. Misalnya saja, gugus data tersebut mungkin mengandung banyak sekali objek amatan yang aneh dan keberadaannya tidak dapat dijelaskan secara sederhana. Amatan seperti itu sering disebut sebagai amatan berpengaruh atau pencilan. Menentukan apakah sebuah amatan berpengaruh atau tidak adalah sesuatu yang cukup sulit. Pendekatan formal dan grafik dapat dilihat pada berbagai literatur.

Situasi lain dimana analisis sederhana yang hanya mengandalkan rata-rata dapat jadi tidak mencukupi adalah ketika data pada beberapa peubah saling berkorelasi dan/atau berpola. Situasi seperti ini sering

muncul pada pengamatan deret waktu. Misalnya, jika data dikumpulkan dari seorang atau sekelompok pasien dengan perlakuan tertentu, kita jarang tertarik mengetahui rata-rata respon dari waktu ke waktu. Yang lebih menarik adalah mengamati perubahan nilainya, yaitu mengamati pola atau trend-nya.

Tidak jarang, data dikumpulkan dari sejumlah unit objek, dan di setiap objek tidak hanya satu amatan, tapi banyak peubah yang diukur. Sebagai contoh, dalam percobaan psikologi, banyak pengujian dilakukan terhadap setiap individu. Karena pengukuran-pengukuran (melalui tes) dilakukan pada individu yang sama, maka pengukuran-pengukuran tersebut saling berkorelasi dan pada saat melakukan peringkasan terhadap data maka informasi korelasi ini harus menjadi bagian yang tak terpisahkan dari hasil ringkasan. Lebih jauh, ketika banyak sekali peubah yang terlibat untuk mendapatkan informasi yang lebih jelas dan lebih mudah, ringkasan tentang korelasi ini mutlak diperlukan. Serta masih banyak lagi hal-hal yng membuat data menjadi lebih kompleks untuk di analisis.

Secara umum, kita memiliki *n* buah amatan dan di setiap amatan dilakukan pengukuran *p* buah karakteristik (peubah), katakanlah x_1, x_2, …, x_p. Selanjutnya data tersebut dapat digambarkan sebagai matriks *n x p*

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}$$

Tentu saja, pengukuran pada baris ke-*i*, yaitu x_{i1}, x_{i2}, …, x_{ip} yang merupakan pengukuran pada individu yang sama, saling berkorelasi. Jika kita menyusun pengamatan tersebut sebagai vektor baris $\mathbf{x}_i$ diperoleh

$$\mathbf{x}_i = [x_{i1}, x_{i2}, \ldots, x_{ip}]'$$

maka $\mathbf{x}_i$ dapat disebut sebagai pengamatan peubah ganda. Dengan demikian, *n* baris pada matriks **X** berpadanan dengan *n* buah pengamatan peubah ganda, dan pengamatan pada setiap $\mathbf{x}_i$ umumnya saling berkorelasi. Sementara itu, antara $\mathbf{x}_1$, $\mathbf{x}_2$, ..., $\mathbf{x}_n$ dapat berkorelasi, dapat juga tidak. Umumnya diasumsikan bahwa antar amatan $\mathbf{x}_1$, $\mathbf{x}_2$, ..., $\mathbf{x}_n$ tidak berkorelasi (saling bebas sebagai asumsi yang lebih kuat), namun ini tidak selalu terjadi. Misalnya saja jika $\mathbf{x}_i$ adalah pengukuran tinggi dan berat anak ke-i dari sebuah keluarga *n* orang anak, maka sangat masuk akal jika antar baris pada matriks **X** memiliki korelasi.

Pada buku ini, kita tidak melakukan pembahasan untuk kasus adanya korelasi antar baris matriks **X**, dengan kata lain diasumsikan baris-baris matriks **X** berpadanan dengan satuan contoh yang diambil secara bebas stokastik.

2.2. Populasi versus Contoh

Seperti telah dibahas sebelumnya, *n* buah baris pada matriks **X** menggambarkan *n* buah amatan peubah ganda. Jika gugus dari *n* buah ini berpadanan dengan keseluruhan unit yang mungkin, maka data yang tersedia adalah data populasi. Sebagai teladan, data dikumpulkan dari seluruh kota di Bogor yang memiliki populasi satu juta atau lebih orang, dan diukur tiga peubah misalnya biaya hidup, rata-rata gaji per tahun, dan kualitas fasilitas kesehatan. Karena setiap warga Bogor yang memenuhi definisi termasuk di dalam data tersebut, maka hasil ringkasan data itu akan menjadi ringkasan populasi yang sesungguhnya.

Tetapi lebih sering data diperoleh dari hasil survei dan dari setiap objek survei diamati *p* buah peubah. Situasi ini menggambarkan contoh peubah ganda. Sebuah contoh (baik cukup atau tidak) merupakan

perwakilan dari populasi tertentu. Karena populasi hanya diwakili oleh sebagian unit yang terpilih, maka hasil ringkasan terhadapnya diharapkan mewakili nilai yang sebenarnya dan diharapkan mendekati nilai tersebut, meskipun tidak ada jaminan yang pasti.

Bagaimana caranya mengukur dan memastikan bahwa hasil ringkasan dari contoh akan merupakan perwakilan yang baik bagi ringkasan populasi? Untuk hal tersebut, berbagai jenis indeks berbasis konsep peluang digunakan. Hal ini mengharuskan seseorang untuk membangun suatu struktur peluang pada setiap amatan. Ide ini diimplementasikan dengan memasukkan struktur peluang pada skema pengambilan contoh, baik secara artifisial maupun sesungguhnya. Tentu saja, jika kita ingin memastika contoh kita adalah wakil yang baik bagi populasi, maka struktur peluang yang dibuat adalah sedemikian rupa sehingga perlakuan terhadap seluruh unit populasi sama (*equally fair*). Sehingga, kita perlu melakukan pengambilan contoh dengan memberikan peluang yang sama bagi setiap unit untuk terpilih sebagai unit contoh. Persyaratan ini dapat terpenuhi dengan melakukan pengambilan contoh secara acak.

Meskipun struktur peluang diperkenalkan terhadap unit penagamatan melalui penarikan contoh acak, tidak demikian halnya dengan p buah pengukuran. Antar pengukuran itu sendiri memungkinkan adanya keterkaitan, sehingga muncul istilah sebaran peluang bersama. Jadi, kita memandang baris-baris matriks $\mathbf{X}$ sebagai pengamatan peubah ganda dari populasi berdimensi-p yang mewakili sebaran peubah ganda berdimensi-p. Dengan kata lain, baris-baris $\mathbf{X}$ adalah contoh acak dari populasi berdimensi-p. Dalam banyak analisis, seringkali populasi diasumsikan tidak terhingga dan diasumsikan memlikiki sebaran normal ganda.

2.3. Beberapa alat dasar untuk memahami data peubah ganda

Untuk memahami data yang besar dan peubah-peubahnya tidak saling bebas, peringkasan tetap harus dilakukan. Untuk data *univariate* (satu peubah yang menjadi fokus pembahasan), peringkasan umumnya dilakukan menggunakan rata-rata, ragam, kemenjuluran, dan kurtosis, baik untuk populasi maupun contoh. Data peubah ganda juga memiliki hal yang serupa. Pada tulisan ini, notasi matriks akan banyak digunakan untuk menyederhanakan penulisan. Beberapa istilah matriks akan dibahas kemudian.

Misalnya **x** adalah vektor acak berukuran *p x 1* yang berpadanan dengan sebuah populasi peubah ganda, atau

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{bmatrix}$$

maka setiap x_i adalah peubah acak, dan kita mengasumsikan $x_1, \ldots, x_p$ mungkin saling tidak bebas. Notasi E(·) menunjukkan nilai harapan (diinterpretasikan sebagai rata-rata dalam jangka panjang), dan misalkan $\mu_i = E(x_i)$ dan $\sigma_{ii} = var(x_i)$ adalah ragam populasi. Selanjutnya peragam populasi antara x_i dan x_j adalah $\sigma_{ij} = cov(x_i, x_j)$. Didefinisikan vektor rataan populasi (**μ**) sebagai vektor dari nilai harapan setiap peubah, yaitu:

$$E(\mathbf{x}) = \begin{bmatrix} E(x_1) \\ \vdots \\ E(x_p) \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_p \end{bmatrix} = \boldsymbol{\mu}$$

Sebagai tambahan, konsep ragam populasi dirangkum dalam sebuah matriks yang memuat ragam dan peragam populasi yang diletakkan bersesuasian dalam matriks ragam-peragam. Jika matriks tersebut dilambangkan Σ, maka

$$\Sigma = \begin{bmatrix} \text{var}(x_1) & \text{cov}(x_1, x_2) & \cdots & \text{cov}(x_1, x_p) \\ \text{cov}(x_2, x_1) & \text{var}(x_2) & \cdots & \text{cov}(x_2, x_p) \\ \vdots & \vdots & \ddots & \vdots \\ \text{cov}(x_p, x_1) & \text{cov}(x_p, x_2) & \cdots & \text{var}(x_p) \end{bmatrix} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp} \end{bmatrix}$$

Dengan mengartikan $\text{cov}(x_i, x_i) = \text{var}(x_i) = \sigma_{ii}$, bentuk $\text{cov}(x_i, x_j)$ dapat disebut sebagai unsur ke-*(i, j)* dari matirks Σ. Nilai-nilai ragam peubah ke-*i* ditempatkan pada diagonal utama ke-*i*, dan nilai peragam akan ditempatkan bersesuaian pada unsur non-diagonal. Karena $\text{cov}(x_i, x_j) = \text{cov}(x_j, x_i)$ atau $\sigma_{ij} = \sigma_{ji}$, maka Σ merupakan matriks yang simetrik.

Nilai $\text{tr}(\Sigma) = \sum_{i=1}^{p} \sigma_{ii}$ disebut sebagai ragam total dan $\det(\Sigma) = |\Sigma|$ disebut sebagai ragam terampat (generalized variance). Dua buah nilai tersebut seringkali digunakan sebagai ukuran keragaman total dari vektor acak **x**. Namun demikian, kadang-kadang penggunaannya dapat menyesatkan, sebgai misal, $\text{tr}(\Sigma)$ sama sekali tidak memperhitungkan nilai-nilai selain diagonal utama yang menunjukkan peragam antar peubah. Atau dapat juga, dua buah matriks yang sangat berbeda mungkin memiliki determinan yang sama.

Karena ada keterkaitan antara x_i, ..., x_p maka masih relevan jika kita melihat tingkat keterkaitannya, palingtidak keterkaitan linearnya melalui besarnya korelasi. Koefisien korelasi Pearson untuk populasi antara x_i dan x_j diperoleh melalui

$$\rho_{ij} = \frac{\mathrm{cov}(x_i, x_j)}{\sqrt{\mathrm{var}(x_i)\,\mathrm{var}(x_j)}} = \frac{\sigma_{ij}}{\sqrt{\sigma_{ii}\sigma_{jj}}}$$

Selanjutnya kita definisikan matriks korelasi populasi sebagai

$$\rho = \begin{bmatrix} \rho_{11} & \rho_{12} & \cdots & \rho_{1p} \\ \rho_{21} & \rho_{22} & \cdots & \rho_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{p1} & \rho_{p2} & \cdots & \rho_{pp} \end{bmatrix} = \begin{bmatrix} 1 & \rho_{12} & \cdots & \rho_{1p} \\ \rho_{21} & 1 & \cdots & \rho_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{p1} & \rho_{p2} & \cdots & 1 \end{bmatrix}$$

Seperti halnya **Σ**, **ρ** juga meupakan matriks simetirk. Lebih jauh, **ρ** dapat dituliskan dalam **Σ** sebagai

$$\rho = \left(diag(\mathbf{\Sigma})\right)^{-\frac{1}{2}} \mathbf{\Sigma} \left(diag(\mathbf{\Sigma})\right)^{-\frac{1}{2}}$$

dengan diag(Σ) adalah matriks diagonal yang didapatkan dengan mempertahankan unsur diagonal Σ dan mengganti unsur non-diagonalnya dengan 0, dan akar kuadrat dari matriks **A** dinotasikan $\mathbf{A}^{1/2}$ adalah matriks yang memenuhi **A** = $\mathbf{A}^{1/2}$ $\mathbf{A}^{1/2}$, dan $\mathbf{A}^{-1/2}$ adalah invers (kebalikan) dari matriks $\mathbf{A}^{1/2}$**.** Pada buku ini diasumsikan matriks ragam-peragam dan matrik korelasi bersifat definit positif.

Bagaimana kita mengukur kemenjuluran (*skewness*) dan kurtosis untuk populasi peubah ganda? Mardia (1970) mendefinisikan ukuran ini sebagai:

multivariate skewness : $\beta_{1,p} = E\left[(\mathbf{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu})\right]^3$, dengan **x** dan **y** saling bebas dan dari sebaran yang sama.

kurtosis multivariate : $\beta_{2,p} = E\left[(\mathbf{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right]^2$.

Untuk kasus univariate, yaitu *p* = 1, $\beta_{1,p}$ menjadi kuadrat dari koefisien kemenjuluran, dan $\beta_{2,p}$ adalah koefisien kurtosis.

Besaran-besaran $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$, $\boldsymbol{\rho}$, $\beta_{1,p}$, dan $\beta_{2,p}$ merupakan nilai-nilai ringkasan dasar bagi populasi peubah ganda. Lalu, apa padanan dari besaran-besaran ini untuk contoh? Jika kita memiliki contoh acak berukuran *n* yang terdiri atas *p* buah peubah $\mathbf{x}_1$, …, $\mathbf{x}_n$, maka didefinisikan matriks **X** yang berukuran *n x p*

$$\mathbf{X}_{nxp} = \begin{bmatrix} \mathbf{x}_1' \\ \vdots \\ \mathbf{x}_n' \end{bmatrix},$$

padanan besaran di atas adalah:

vektor rataan contoh : $\bar{\mathbf{x}} = n^{-1}\sum_{i=1}^{n}\mathbf{x}_i = n^{-1}\mathbf{X}'\mathbf{1}_n$

matriks ragam-peragam contoh : $\mathbf{S} = (n-1)^{-1}\sum_{i=1}^{n}(\mathbf{x}_i-\bar{\mathbf{x}})(\mathbf{x}_i-\bar{\mathbf{x}})'$

$$= (n-1)^{-1}\left\{\sum_{i=1}^{n}\mathbf{x}_i\mathbf{x}_i' - n\bar{\mathbf{x}}\bar{\mathbf{x}}'\right\}$$

$$= (n-1)^{-1}\left[\mathbf{X}'(\mathbf{I} - n^{-1}\mathbf{1}_n\mathbf{1}_n')\mathbf{X}\right]$$

$$= (n-1)^{-1}\left[\mathbf{X}'\mathbf{X} - n^{-1}\mathbf{X}'\mathbf{1}_n\mathbf{1}_n'\mathbf{X}\right]$$

$$= (n-1)^{-1}\left[\mathbf{X}'\mathbf{X} - n\bar{\mathbf{x}}\bar{\mathbf{x}}'\right]$$

Seringkali juga disebutkan dalam beberapa literatur bahwa pembagi pada *persamaan* di atas adalah *n* bukan *(n – 1)*. Jika demikian halnya maka notasi yang digunakan adalah $\mathbf{S}_n$. Kita juga masih memiliki beberapa besaran, yaitu:

matriks korelasi contoh : $\hat{\boldsymbol{\rho}} = \left[diag(\mathbf{S})\right]^{-\frac{1}{2}}\mathbf{S}\left[diag(\mathbf{S})\right]^{-\frac{1}{2}}$

$$= \left[diag(\mathbf{S}_n)\right]^{-\frac{1}{2}}\mathbf{S}_n\left[diag(\mathbf{S}_n)\right]^{-\frac{1}{2}}$$

skewness contoh peubah ganda : $\hat{\beta}_{1,p} = n^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n} g_{ij}^3$,

dan

kurtosis contoh peubah ganda : $\hat{\beta}_{2,p} = n^{-2}\sum_{i=1}^{n} g_{ii}^2$.

Pada persamaan-persamaan di atas, $\mathbf{1}_n$ adalah vektor kolom berukuran *n x 1* dengan seluruh unsur bernilai 1, $\mathbf{I}_n$ adalah matriks identitas berukuran *n x n*, dan g_{ij}, *i, j* = 1, 2, ..., *p*, didefinisikan sebagai g_{ij} = $(\mathbf{x}_i - \bar{\mathbf{x}})'\mathbf{S}_n^{-1}(\mathbf{x}_j - \bar{\mathbf{x}})$.

2.4. Pereduksian Data, Pendeskripsian, dan Pendugaan

Pada bagian sebelumnya telah dibahas beberapa besaran penciri dasar populasi dan padanannya untuk contoh, yang biasa disebut statistik deskriptif. Ide dasarnya adalah merangkum data populasi atau data contoh menjadi matriks yang lebih kecil ukurannya atau menggunakan angka-angka sederhana. Semua besaran-besaran tadi (kecuali korelasi) hanya merupakan padanan besaran yang sama pada kasus univariate. Namun demikian, data peubah ganda memiliki ciri dan kebutuhan yang khas, yang tidak ada pada situasi univariate. Meskipun ide dasarnya tetap sama, yaitu meringkas atau mendeskripsikan data, situasi tertentu mungkin memerlukan teknik yang khas. Beberapa teladan untuk itu antara lain:

a. Berdasarkan beberapa peubah (variabel) seperti rata-rata harga rumah, biaya hidup, kelengkapan fasilitas kesehatan, tingkat kriminalitas, dan sebagainya, kita ingin mendeskripsikan kota-kota mana saja yang sesuai untuk kelayakan hidup tertentu dan mengamati adanya kesamaan dan perbedaan antar kota. Ada beberapa peubah yang diukur, dan antar peubah mungkin saling bertolak belakang. Sebagai misal, kota yang rendah tingkat kejahatannya (karakteristik yang diinginkan) cenderung memiliki biaya hidup yang tinggi (karakteristik yang tidak diinginkan), jadi keduanya saling bertolak belakang. Bagaimana kita menentukan kota mana yang paling baik untuk ditinggali? Ini adalah masalah pereduksian data. Masalah ini tidak dapat dipandang sebagai penyeleksian peubah karena tidak ada peubah tak bebasnya dan tidak ada model, serta tidak dapat dipandang sebagai masalah pendugaan. Permasalahan ini lebih mendekati upaya pendeteksian dan pemahaman karakteristik yang terkandung pada data, dan kemudian menginterpretasikannya. Situasi ini memerlukan beberapa pendekatan untuk mendeskripsikan data. Analisis yang mungkin

dapat digunakan adalah analisis komponen utama atau analisis gerombol.

b. Misalkan kita memiliki beberapa peubah bebas yang dicurigai memiliki pengaruh terhadap beberapa (banyak) peubah tak bebas. Situasi seperti ini seringkali ditemui pada bidang industri dan ekonomi, dimana dua gugus peubah dapat didefinisikan secara tegas dan memuat peubah input dan output. Kita tidak tertarik pada masing-masing peubah, tapi kita ingin memperoleh lebih sedikit peubah baru pada setiap kelompok peubah. Peubah baru ini mungkin saja dalah fungsi dari semua peubah asal pada masing-masing kelompok, dan berharap peubah baru ini dapat diinterpretasikan dengan tepat. Kita harus menekankan bahwa analisis ini tidak dilakukan dengan tujuan untuk membuktikan atau menyanggah pernyataan tertentu. Ini hanyalah upaya untuk memahami data. Karena informasi telah terangkum pada pubah yang baru, dan jumlahnya lebih sedikit, maka keterkaitan yang ada diharapkan lebih mudah untuk dilihat. Permasalahan ini antara lain dapat ditangani oleh analisis korelasi kanonik, dan analisis korespondensi untuk data kualitatif (kategorik).

c. Sebuah perusahaan mobil ingin mengetahui apa yang menentukan pilihan konsumen pada berbagai mobil. Contoh acak 100 orang yang terpilih diminta untuk memberikan skor antara 1 (rendah) hingga 10 (tinggi) pada 6 peubah, yaitu harga, ketahanan, simbol status bekenaan dengan mobil, konsumsi bahan bakar, keamaan ketika kecelakaan, dan rata-rata jarak tempuh per minggu. Analsis jenis apa yang dapat digunakan untuk hal seperti ini? Dengan mengasumsikan adanya peubah hipotetik dan tak teramati yang mempengaruhi skor dari keenam peubah tadi, pertanyaan yang muncul adalah peubah hipotetik apakah itu. Secara intuitif, keamanan dan sisi status ekonomi dapat menjadi dua hal yang mempengaruhi skor-skor tadi. Jadi, beberapa atau semua peubah

yang teramati dapat dituliskan sebagai fungsi dari dua peubah hipotetik. Atau mungkin pertanyaan dibalik, apakah peubah yang tak teramati merupakan fungsi dari peubah yang teramati. Analisis faktor dapat dijadikan pilihan untuk kasus yang demikian. Sebagai catatan, analisis ini hanya memberikan fungsi, pemaknaan terhadap peubah yang tak teramati ditinggalkan sebagai bagian untuk analis. Jadi, permasalahan ini juga merupakan pereduksian dari banyak peubah menjadi sedikit peubah.

2.5. Konsep-Konsep Aljabar Matriks

Sub-bab ini dimaksudkan hanya sebagai ulasan singkat beberapa konsep aljabar matriks. Kami mengasumsikan bahwa para pembaca sudah tidak asing dengan operasi penjumlahan, penggandaan, dan pemutaran matriks. Serta beberapa konsep dasar seperti kebebasan linear.

2.5.1. Matriks Putaran

Untuk matriks **A** berukuran *m x n*, putaran dari matriks **A** diperoleh dengan cara menukar baris dan kolomnya, dan dinotasikan **A**' berukuran *n x m*. Dalam notasi jika $\mathbf{A} = (a_{ij})$ dan putarannya adalah $\mathbf{A}' = (a'_{kl})$ maka $a_{ij} = a'_{ji}$. Sebagai teladan, jika

$$\mathbf{A} = \begin{bmatrix} 1 & 3 & 4 \\ 7 & 0 & 1 \end{bmatrix} \text{ maka } \mathbf{A}' = \begin{bmatrix} 1 & 7 \\ 3 & 0 \\ 4 & 1 \end{bmatrix}$$

Juga untuk matriks **A** berukuran *m x n,* dan **B** berukuran *n x r,* berlaku

$$(\mathbf{AB})' = \mathbf{B}'\mathbf{A}'.$$

2.5.2. Matriks Simetrik

Sebuah matriks **A** berukuran *n x n* dikatakan simetrik jika **A'** = **A**. Sebagai teladan

$$\mathbf{A} = \begin{bmatrix} 7 & 8 & -3 \\ 8 & 0 & 1 \\ -3 & 1 & 9 \end{bmatrix}$$

adalah matriks simetrik. Jelasnya, jika a_{ij} adalah unsur ke-*(i, j)* dari matriks **A**, maka untuk matriks simetrik $a_{ij} = a_{ji}$, untuk semua i dan j. Dalam banyak analisis statistik peubah ganda, nantinya kita akan sering berhadapan dengan matriks ragam-peragam, **Σ**, dan matriks korelasi, **ρ**, yang keduanya adalah matriks simetrik karena $\sigma_{ij} = \sigma_{ji}$ dan $\rho_{ij} = \rho_{ji}$.

2.5.3. Matriks Diagonal

Sebuah matriks **A** berukuran *n x n* disebut matriks diagonal jika semua unsur non-diagonalnya bernilai 0. Matriks diagonal tentulah matriks yang simetrik. Contoh matriks diagonal adalah

$$\mathbf{A} = \begin{bmatrix} 7 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 9 \end{bmatrix}$$

Pada situasi tertentu digunakan notasi diag(**A**), yang berarti sebuah matriks yang mempertahankan unsur-unsur diagonal **A** dan mengganti unsur non-diagonal dengan 0. Jadi untuk

$$\mathbf{A} = \begin{bmatrix} 7 & 8 & -3 \\ 8 & 0 & 1 \\ -3 & 1 & 9 \end{bmatrix}, \text{ maka diag}(\mathbf{A}) = \begin{bmatrix} 7 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 9 \end{bmatrix}.$$

2.5.4. Beberapa Matriks Khusus

Terdapat beberapa matriks yang memiliki ciri-ciri tertentu dilihat dari unsur dari matriks tersebut, antara lain:

a. Matriks diagonal berukuran *n x n* dengan semua unsur diagonalnya bernilai 1 dan unsur non-diagonalnya bernilai 0 disebut sebagai matriks identitas. Notasi yang digunakan adalah $\mathbf{I}_n$ atau disingkat **I** jika sudah jelas ukurannya.

b. Matriks berukuran *m x n* dengan seluruh unsurnya bernilai 0 disebut matriks nol. Biasanya dinotasikan $\mathbf{0}_{m,n}$ atau disingkat **0**.

2.5.5. Matriks Segitiga

Sebuah matriks **A** berukuran *n x n* disebut matriks segitiga atas jika semua unsur di bawah diagonal utama bernilai 0. Sedangkan matriks diagonal bawah adalah matriks yang semua unsur di atas diagonal utama bernilai 0. Sebagai contoh,

$\mathbf{A}_1 = \begin{bmatrix} 1 & 1 & 9 \\ 0 & 3 & 4 \\ 0 & 0 & 4 \end{bmatrix}$ adalah matriks segitiga atas, sedangkan $\mathbf{A}_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 3 & 9 \end{bmatrix}$ adalah matriks segitiga bawah. Dalam banyak operasi aljabar, pelibatan matriks segitiga akan mempermudah proses perhitungan (seperti pencarian determinan, akar ciri, penyelesaian sistem persamaan linear) sehingga mempercepat proses numerik yang dilakukan.

2.5.6. Kebebasan Linear

Segugus vektor kolom (atau baris) tak nol dikatakan bebas linear jika tidak ada satupun yang dapat dituliskan sebagai kombinasi linear dari beberapa atau semua vektor lainnya. Jika hal ini tidak terjadi maka gugus tersebut dikatakan tak bebas linear. Gugus yang mengandung vektor nol selalu merupakan gugus tak bebas linear.

Dalam notasi, segugus vektor S = {s1, …, sm} dikatakan gugus yang bebas linear jika untuk semua $\mathbf{s}_i$, $i = 1, \ldots, m$, tidak ada yang memenuhi

$$\mathbf{s}_j = \sum_{i \neq j} a_i \mathbf{s}_i \text{ ; } a_i \text{ adalah konstanta}$$

Dapat juga ditunjukkan bahwa ungkapan di atas setara dengan mengatakan $S = \{\mathbf{s}_1, \ldots, \mathbf{s}_m\}$ dikatakan gugus yang bebas linear jika

$$\sum_{i=1}^{m} a_i \mathbf{s}_i = 0$$

terjadi hanya jika semua $a_i = 0$, $i = 1, \ldots, m$.

Jika ada gugus vektor yang tak bebas linear, dan kita membuang vektor nol, kemudian melanjutkan dengan membuang vektor-vektor yang dapat dinyatakan sebagai kombinasi vektor lainnya, maka kita akan mendapatkan gugus yang bebas linear atau mungkin juga kita dapatkan gugus kosong. Banyaknya vektor pada hasil akhir yang tersisa tadi merupakan salah satu konsep penting yang akan dibahas selanjutnya.

2.5.7. Pangkat Matriks

Pangkat dari sebuah matriks **A**, dilambangkan r(**A**) didefinisikan sebagai banyaknya baris (atau kolom) pada matriks itu yang bersifat bebas linear. Dalam hal ini kita dapat bekerja pada baris atau kolom, sehingga jelas bahwa r(**A**) = r(**A'**). Lebih jauh, dapat juga ditunjukkan bahwa r(**AB**) ≤ min { r(**A**), r(**B**) }, dan r(**A'A**) = r(**A**).

2.5.8. Matriks Non-Singular dan Matriks Singular

Sebuah matriks **A** berukuran *n x n* dikatakan non-singular jika semua baris (atau kolom)-nya saling bebas linear. Dengan kata lain, **A** berukuran *n x n* non-singular jika r(**A**) = *n*. Jika satu atau lebih baris (atau kolom) **A** dapat dituliskan sebagai kombinasi linear dari beberapa atau semua baris (atau kolom) lainnya dari **A**, maka ada ketakbebasan linear diantaranya. Pada kasus ini, **A** dikatakan singular. Sebagai teladan

$$\mathbf{A} = \begin{bmatrix} 1 & 3 \\ 9 & 4 \end{bmatrix}$$

adalah non-singular karena tidak satupun baris yang dapat dituliskan dalam bentuk baris yang lain. Tetapi

$$\mathbf{B} = \begin{bmatrix} 1 & 3 & 4 \\ 9 & 4 & 3 \\ 11 & 10 & 11 \end{bmatrix}$$

adalah singular karena $baris_3$ = 2 x $baris_1$ + $baris_2$, yang menunjukkan bahwa baris ke-3 dapat dituliskan dalam bentuk kombinasi dua baris lainnya.

2.5.9. Kebalikan Matriks Persegi

Sebuah matriks **A** berukuran n x n memiliki kebalikan jika ada matriks **B** sehingga **AB** = **BA** = $\mathbf{I}_n$. Matriks **B** disebut sebagai *kebalikan matriks* **A**, dan dinotasikan $\mathbf{A}^{-1}$. Sebagai misal,

$$\mathbf{A} = \begin{bmatrix} 1 & 3 \\ 9 & 4 \end{bmatrix}$$

matriks $\mathbf{A}^{-1}$ adalah

$$\mathbf{A}^{-1} = \begin{bmatrix} -\frac{4}{23} & \frac{3}{23} \\ \frac{9}{23} & -\frac{1}{23} \end{bmatrix}$$

Jelas bahwa kebalikan dari $\mathbf{A}^{-1}$, yaitu $(\mathbf{A}^{-1})^{-1}$, adalah **A**. Matriks kebalikan hanya didefinisikan pada matriks berukuran *n x n*, yaitu jika banyaknya baris dan kolom sama. Untuk matriks demikian, kebalikan ada jika dan hanya jika **A** non-sisngular. Jadi, matriks kebalikan tidak dimiliki oleh matriks yang singular, atau matriks yang banyaknya baris dan kolomnya tidak sama. Untuk matriks seperti itu, ada konsep yang lebih lemah yaitu kebalikan umum kondisional, atau *conditional inverse*.

Jika ada dua buah matriks **A** dan **B** berukuran *n x n* dan non-singular, maka $(\mathbf{AB})^{-1}$ dan $(\mathbf{BA})^{-1}$ ada, walaupun tidak sama. Secara khusus,

$$(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\,\mathbf{A}^{-1}$$

dan

$$(\mathbf{BA})^{-1} = \mathbf{A}^{-1}\,\mathbf{B}^{-1}$$

Karena sifat komutatif tidak berlaku pada penggandaan maka $(\mathbf{AB})^{-1}$ dan $(\mathbf{BA})^{-1}$ tidak sama.

2.5.10. Kebalikan Umum

Untuk sebuah matriks **B** berukuran *m x n*, matriks kebalikan umum kondisional bagi **B**, misalkan **G** adalah matriks berukuran *n x m* yang memenuhi

$$\mathbf{BGB} = \mathbf{B}$$

Secara umum, matriks kebalikan umum kondisional ini selalu ada, namun tidak bersifat unik (khas). Matriks kebalikan umumkondisiona ini bersifat unik untuk matriks non-singular, dan pada kasus ini sama dengan matriks kebalikan biasa. Matriks kebalikan umum bagi **B** dinotasikan dengan $\mathbf{B}^{-}$.

Matriks

$$\mathbf{B} = \begin{bmatrix} 1 & 3 & 4 \\ 9 & 4 & 3 \\ 11 & 10 & 11 \end{bmatrix}$$

sudah ditunjukkan bersifat singular. Matriks kebalikan umum bagi **B** adalah

$$\mathbf{B}^{-} = \begin{bmatrix} -\frac{4}{23} & \frac{3}{23} & 0 \\ \frac{9}{23} & -\frac{1}{23} & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Tentu saja $\mathbf{B}^{-}$ di atas bukan satu-satunya.

2.5.11. Sistem Persamaan Linear

Misalkan sebuah sistem terdidri atas *n* buah persamaan dengan *m* peubah tak diketahui, x_1, …, x_m, yang konsisten, yaitu sistem yang tidak memiliki anak gugus persamaan yang tidak bertentangan dengan persamaan sisanya.

$$a_{11} x_1 + a_{12} x_2 + \ldots + a_{1m} x_m = b_1$$

$$a_{21} x_1 + a_{22} x_2 + \ldots + a_{2m} x_m = b_1$$

.

.

$$a_{n1} x_1 + a_{n2} x_2 + \ldots + a_{nm} x_m = bn$$

Dengan mendefinisikan

$$\mathbf{A} = \begin{bmatrix} a_{11} & \cdots & a_{1m} \\ \vdots & & \\ a_{n1} & \cdots & a_{nm} \end{bmatrix}, \mathbf{b} = \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}, \text{ dan } \mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_m \end{bmatrix},$$

persamaan di atas dapat dituliskan sebagai berikut:

$$\mathbf{Ax} = \mathbf{b}$$

Jika $m = n$ dan matriks **A** non-singular, maka solusi untuk **x** diberikan oleh $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$. Jika $m = n$ dan jika **A** bersifat singular (pada kasus dimana beberapa persamaan redundan karena sudah tercakup pada persamaan-persamaan lain) atau jika $m \neq n$, maka solusi bagi **x** adalah $\mathbf{x} = \mathbf{A}^{-}\mathbf{b}$, dimana $\mathbf{A}^{-}$ adalah matriks kebalikan umum kondisional bagi **A**. Karena matriks kebalikan umum kondisional tidak unik kecuali pada **A** yang non-singular, maka pada kasus ini tidak diperoleh solusi yang unik untuk sistem persamaan tersebut. Alasannya adalah dengan memilih

matriks kebalikan umum yang lain maka akan diperoleh solusi yang berbeda pula.

Ketika **A** non-singular, banyaknya solusi tak hingga, dan kesemuanya terangkum pada persamaan **x** = $\mathbf{A}^{-}\mathbf{b}$ + ($\mathbf{I}$ – $\mathbf{A}^{-}\mathbf{A}$)**z**, dengan **z** adalah sembarang vektor dan $\mathbf{A}^{-}$ adalah matriks kebalikan umum kondisional bagi **A**. Bentuk khusus dari sistem persamaan linear adalah sistem persamaan linear homogen, **Ax** = **0**. Meskipun dapat jadi banyaknya solusi tak hingga, namun pada kasus sistem yang homogen dapat dicari segugus solusi yang saling ortonormal (akan didefinisikan kemudian), sebut saja dikumpulkan pada matriks **X**. Kolom-kolom matriks ini adalah solusi yang ortonormal, dan ordo matriks ini ditentukan oleh pangkat matriks **A**.

2.5.12. Norma (Panjang) Vektor Euclidean

Untuk sebuah vektor **a** berukuran *n x 1*, norma (atau panjang) dari **a** didefinisikan sebagai $\sqrt{\mathbf{a}'\mathbf{a}}$. Jelas bahwa vektor **b** yang didefinisikan sebagai **b** = **a**/ $\sqrt{\mathbf{a}'\mathbf{a}}$ memiliki norma 1. Pada kasus ini **b** dikenal sebagai vektor **a** yang *dinormalkan*. Bentuk $\sqrt{\mathbf{a}'\mathbf{a}}$ tidak lain adalah sama dengan $\sum_{i=1}^{n} a_i^2$, dengan a_i (*i = 1, …, n*) adalah unsur ke-*i* dari vektor **a**.

Sebagai contoh, vektor

$$\mathbf{a} = \begin{bmatrix} 2 \\ 1 \\ 2 \end{bmatrix}$$

memiliki norma sebesar $\sqrt{2^2 + 1^2 + 2^2}$ = 3, dan vektor

$$\mathbf{b} = \frac{1}{3}\begin{bmatrix} 2 \\ 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 2/3 \\ 1/3 \\ 2/3 \end{bmatrix}$$

memiliki norma 1, karena **b** adalah vektor **a** yang dinormalkan.

2.5.13. Jarak Euclid antar Dua Vektor

Dengan menggambarkan vektor **a** dan **b** berukuran *n x 1* sebagai titik pada ruang berdimensi-*n*, kita dapat mendefinisikan jarak antara **a** dan **b** sebagai norma dari vektor (**a – b**). dengan demikian, jarak *d*(**a, b**) didefinisikan sebagai

$$d(\mathbf{a}, \mathbf{b}) = \sqrt{(\mathbf{a}-\mathbf{b})'(\mathbf{a}-\mathbf{b})}$$

$$= \sqrt{\sum_{i=1}^{n}(a_i - b_i)^2} \text{ ,}$$

dengan a_i dan b_i berurutan adalah unsur ke-*i* dari vektor **a** dan **b**.

Sebagai contoh dua buah vektor

$$\mathbf{a} = \begin{bmatrix} 5 \\ 3 \\ 2 \end{bmatrix} \text{ dan } \mathbf{b} = \begin{bmatrix} 6 \\ 1 \\ 4 \end{bmatrix}$$

memiliki jarak Euclid sebesar

$$d(\mathbf{a}, \mathbf{b}) = \sqrt{(5-6)^2 + (3-1)^2 + (2-4)^2} = 3$$

Jarak Euclid adalah jarak antar titik seperti yang kita lihat menggunakan mata. Namun demikian, kadangkala beberapa jarak dapat didefinisikan dengan memberikan bobot melalui sebuah matriks definit positif (akan dibahas berikutnya). Jarak yang dibangun dengan memberikan bobot menggunakan matriks pembobot **A** didefinisikan sebagai berikut:

$$d_{\mathbf{A}}(\mathbf{a}, \mathbf{b}) = \sqrt{(\mathbf{a}-\mathbf{b})'\mathbf{A}(\mathbf{a}-\mathbf{b})}$$

Jelas bahwa $d_{\mathbf{I}}(\mathbf{a}, \mathbf{b}) = d(\mathbf{a}, \mathbf{b})$. Salah satu jarak terboboti yang umum digunakan dalam analisis peubah ganda adalah jarak Mahalanobis.

Secara umum, fungsi dari jarak misalnya $d(\mathbf{a}, \mathbf{b})$ dapat didefinisikan dengan banyak cara. Namun fungsi jarak itu haruslah memenuhi persyaratan berikut:

c. $d(\mathbf{a}, \mathbf{b}) = 0$, jika dan hanya jika $\mathbf{a} = \mathbf{b}$
d. $d(\mathbf{a}, \mathbf{b}) = d(\mathbf{b}, \mathbf{a})$
e. $d(\mathbf{a}, \mathbf{b}) \geq 0$
f. $d(\mathbf{a}, \mathbf{c}) \leq d(\mathbf{a}, \mathbf{b}) + d(\mathbf{b}, \mathbf{c})$

Dapat diperiksa bahwa $d(\mathbf{a}, \mathbf{b})$ dan $d_{\mathbf{A}}(\mathbf{a}, \mathbf{b})$ memenuhi syarat di atas. Perlu menjadi catatan bahwa dalam statistika seringkali jarak kuadrat juga disebutkan sebagai jarak. Ini seringkali terjadi pada kasus analisis gerombol. Dalam konteks ini, seringkali juga kita menggunakan jarak sebagai ukuran ketakmiripan (*dissimilarity*) antar objek atau individu. Walaupun, sebenarnya banyak juga indeks ketakmiripan lain yang digunakan, dan indeks ketakmiripan ini tidak selalu memenuhi syarat sebagai jarak.

2.5.14. Vektor dan Matriks Ortogonal

Dua buah vektor berukuran *n x 1*, **a** dan **b** dikatakan ortogonal satu sama lain jika **a'b** = 0. Lebih jauh, jika **a** dan **b** adalah vektor yang dinormalkan (yaitu **a'a** = 1 = **b'b**) maka keduanya disebut ortonormal. Sebagai contoh,

$$\mathbf{a} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \text{ dan } \mathbf{b} = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$$

adalah dua matriks yang saling ortogonal. Jika untuk yang dinormalkan, yaitu **a**/√3 dan **b**/√2 maka keduanya bersifat saling ortonormal.

Sebuah matriks **A** berukuran n x n dikatakan sebagai matriks ortogonal jika

$$\mathbf{A'A} = \mathbf{AA'} = \mathbf{I}_n$$

Hal ini secara cukup setara dengan mengatakan bahwa semua baris (atau kolom) matriks **A** bersifat ortonormal satu dengan yang lain. Karena pada matriks ortogonal berlaku **A'A** = **AA'** = $\mathbf{I}_n$, maka **A'** juga berfungsi sebagai kebalikan matriks **A**. Dengan demikian, **A** juga bersifat non-singular, dan jelas bahwa **A'** juga bersifat ortogonal.

Misalkan *m* < *n* dan **A** berukuran *n x m*, sedemikian rupa sehingga semua kolom matriks **A** bersifat ortonormal. Dalam hal ini

$$\mathbf{A'A} = \mathbf{I}_m$$

tetapi tidak berlaku untuk **AA'**. Jika ini yang terjadi, maka **A** disebut sebagai matriks sub-ortogonal.

Matriks A berikut ini adalah contoh matriks orthogonal,

$$\mathbf{A} = \begin{bmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & 0 & \frac{2}{\sqrt{6}} \end{bmatrix}$$

Tetapi matriks A_1 berikut:

$$\mathbf{A}_1 = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} \\ -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} \\ 0 & \frac{2}{\sqrt{6}} \end{bmatrix}$$

adalah matriks sub-ortogonal karena hanya $\mathbf{A_1'A_1 = I_2}$, tetapi $\mathbf{A_1A_1'}$ tidak sama dengan $\mathbf{I_3}$.

Melakukan penggandaan di depan sebuah matriks dengan menggunakan matriks ortogonal, menghasilkan rotasi sumbu. Hal ini seringkali nanti kita jumpai pada konteks analisis komponen utama dan analisis faktor.

2.5.15. Akar Ciri dan Vektor Ciri

Misalkan **A** adalah matriks berukuran n x n. Pasangan-pasangan (λ_1, $\mathbf{x}_1$), …, (λ_n, $\mathbf{x}_n$) dikatakan sebagai pasangan akar ciri dan vektor ciri jika semua (λ_i, $\mathbf{x}_i$) memenuhi persamaan

$$\mathbf{A\ x} = \lambda\ \mathbf{x}$$

Jika $\mathbf{x}_i$ memenuhi hal di atas, maka kelipatan dari $\mathbf{x}_i$ juga memenuhi. Jadi itulah sebabnya sering kita bekerja dengan vektor ciri $\mathbf{x}_i$ yang normanya 1.

Pada kasus tertentu, nilai akar ciri dan unsur vektor ciri dapat berupa bilangan kompleks. Jika ada akar ciri yang bernilai nol, maka ini juga berarti bahwa matriks **A** bersifat singular.

Jika **A** non-singular maka $\mathbf{A}^{-1}$ ada, dan akar ciri dari $\mathbf{A}^{-1}$ adalah

$$\frac{1}{\lambda_1}, \ldots, \frac{1}{\lambda_m}$$

dengan vektor ciri yang berpadanan sama dengan vektor ciri matriks **A**.

Jika **A** adalah matriks segitiga atas atau segitiga bawah maka akar ciri dari matriks **A** tidak lain adalah sama dengan unsur-unsur diagonal matriks tersebut.

Nilai akar ciri mungkin berulang. Jika akar ciri berulang r kali, maka dikatakan bahwa akar ciri tersebut berulang r. Jika **A** bersifat simetrik, maka vektor ciri yang berpadanan dengan akar ciri yang berbeda bersifat ortonormal (setelah dinormalkan). Lebih jauh, vektor ciri yang berpadanan dengan akar ciri yang berulang r tidak harus ortonormal, tapi kita dapat mendapatkan r vektor ciri berbeda yang bersifat ortonormal. Menggabungkan hal tersebut, kita selalu dapat mendapatkan n buah vektor ciri yang ortonormal dari sebuah matriks simetrik.

Jadi, jika sudah diperoleh vektor ciri yang ortonormal, misalkan $\mathbf{x}_1, \ldots, \mathbf{x}_n$, kita memiliki n buah persamaan

$$\begin{array}{c} \mathbf{Ax_1} = \lambda_1\mathbf{x_1} \\ \vdots \\ \vdots \\ \vdots \\ \mathbf{Ax_n} = \lambda_n\mathbf{x_n} \end{array}$$

Menuliskan persamaan tersebut secara berdampingan menghasilkan persamaan matriks

$$(\mathbf{Ax}_1 \mid \mathbf{Ax}_2 \mid \dots \mid \mathbf{Ax}_n) = (\lambda_1\mathbf{x}_1 \mid \lambda_2\mathbf{x}_2 \mid \dots \mid \lambda_n\mathbf{x}_n)$$

atau

$$\mathbf{A}\,(\mathbf{x}_1 \mid \mathbf{x}_2 \mid \dots \mid \mathbf{x}_n) = (\mathbf{x}_1 \mid \mathbf{x}_2 \mid \dots \mid \mathbf{x}_n)\begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}$$

Misalkan $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$ dan $\mathbf{P} = (\mathbf{x}_1 \mid \mathbf{x}_2 \mid \dots \mid \mathbf{x}_n)$. Jelas bahwa Λ adalah matriks diagonal dan **P** adalah matriks ortogonal, karena semua $\mathbf{x}_i$ bersifat ortonormal. Dengan demikian diperoleh

$$\mathbf{AP} = \mathbf{P}\Lambda$$

atau

$$\mathbf{A} = \mathbf{P}\Lambda\mathbf{P}'$$

Kenyataan di atas merupakan penguraian matriks yang penting pada matriks simetrik, seperti yang akan dibahas berikutnya.

2.5.16. Penguraian Spektral dari Sebuah Matriks Simetrik

Misalkan **A** adalah matriks simetrik berukuran n x n. Maka **A** dapat dituliskan sebagai

$$\mathbf{A} = \mathbf{P}\Lambda\mathbf{P}'$$

dengan **P** adalah suatu matriks ortogonal dan Λ adalah matriks diagonal. Tentu saja pemilihan **P** dan Λ seperti yang telah dibahas sebelumnya. Penguraian seperti ini disebut sebagai penguraian spektral matirks **A.** Karena ketidakunikan vektor ciri, maka penguraian ini pun tidak unik.

2.5.17. Akar Ciri dan Vektor Ciri Umum

Misalkan **A** dan **B** adalah dua matriks simetrik berukuran n x n, dan **B** bersifat definit positif. Maka (δ_1, $\mathbf{x}_1$), …, (δ_n, $\mathbf{x}_n$) adalah pasangan akar ciri dan vektor ciri matriks **A** engan memperhitungkan matriks **B** jika memenuhi persamaan ciri umum

$$\mathbf{A}\, x_i = \delta_i\, \mathbf{B}\, x_i$$

untuk semua i = 1, …, n. Dengan mendefinisikan $\mathbf{Q} = (\mathbf{x}_1 \mid \mathbf{x}_2 \mid \dots \mid \mathbf{x}_n)$, kesemua n persamaan di atas dapat dituliskan dalam persamaan matriks menjadi

$$\mathbf{AQ = BQ\Delta}$$

dimana $\Delta = \mathrm{diag}(\delta_1, \dots, \delta_n)$. Masalah akar ciri umum ini kadangkala muncul pada banyak analisis statistik. Salah satunya adalah pada penyusunan fungsi diskriminan kanonik.

2.5.18. Determinan Matriks

Pada tulisan ini, didefinisikan determinan dari matriks **A** berukuran nxn adalah perkalian dari semua akar ciri **A,** $\lambda_1, \dots, \lambda_n$, dan dinotasikan $|\mathbf{A}|$, sehingga

$$|\mathbf{A}| = \lambda_1 \times \dots \times \lambda_n$$

Jadi |**A**| = 0 jika dan hanya jika paling tidak ada satu akar ciri yang 0, yaitu terjadi jika dan hanya jika **A** singular.

Jika **A** dan **B** adalah dua matriks persegi berukuran n, maka |**AB**| = |**A**|x|**B**| = |**B**|x|**A**| = |**BA**|. Dengan demikian jelas bahwa jika **A** non singular maka $|\mathbf{A}^{-1}|=|\mathbf{A}|^{-1}$, karena |**I**| = 1. Dapat pula ditunjukkan bahwa jika **A** matriks ortogonal maka |**AA'**| = |**I**| = 1 = |**A**||**A'**| = |**A**||**A**| = $|\mathbf{A}|^2$, sehingga |**A**|=1 atau |**A**|= -1.

2.5.19. Teras Matriks

Teras dari matriks **A** berukuran *n x n* didefinisikan sebagai penjumlahan semua akar cirinya, dan dinotasikan *tr*(**A**), sehingga

$$tr(\mathbf{A}) = \lambda_1 + \ldots + \lambda_n$$

Bisa ditunjukkan bahwa $tr(\mathbf{A}) = a_{11} + a_{22} + \ldots + a_{nn}$, jumlah dari unsur-unsur diagonalnya. Padanan ini merupakan konsep yang sangat berguna pada pengembangan teori komponen utama.

Sebagai contoh, jika

$$\mathbf{A} = \begin{bmatrix} 3 & 4 & 1 \\ 2 & 3 & 0 \\ 1 & 2 & 1 \end{bmatrix} \text{ dan } \mathbf{B} = \begin{bmatrix} 2 & 1 & 1 \\ 0 & 1 & 2 \\ 1 & 0 & 2 \end{bmatrix}$$

maka *tr*(**A**) = 3 + 3 + 1 = 7 dan *tri*(**B**) = 2 + 1 + 2 = 5

Jika **A** dan **B** adalah dua buah matriks yang berturut-turut berukuran *mxn* dan *nxm* maka *tr*(**AB**) = *tr*(**BA**). Pada ilustrasi sebelumnya,

$$\mathbf{AB} = \begin{bmatrix} 7 & 7 & 13 \\ 4 & 5 & 8 \\ 3 & 3 & 7 \end{bmatrix} \text{ dan } \mathbf{BA} = \begin{bmatrix} 9 & 13 & 3 \\ 4 & 7 & 2 \\ 5 & 8 & 3 \end{bmatrix}$$

serta $tr(\mathbf{AB}) = 7 + 5 + 7 = 19 = 9 + 7 + 3 = tr(\mathbf{BA})$

2.5.20. Pengutamaan

Misalkan vektor **a** dan **b** didefinisikan seperti berikut,

$$\mathbf{a} = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} \text{ dan } \mathbf{b} = \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}$$

adalah dua vektor berukuran n x 1 dengan $a_1 \geq a_2 \geq \ldots \geq a_n$ dan $b_1 \geq b_2 \geq \ldots \geq b_n$. Maka **a** dikatakan diutamakan oleh **b** jika

$$a_1 \leq b_1$$

$$a_1 + a_2 \leq b_1 + b_2$$

$$a_1 + a_2 + \ldots + a_{n-1} \leq b_1 + b_2 + \ldots + b_n$$

$$a_1 + a_2 + \ldots + a_n = b_1 + b_2 + \ldots + b_n$$

Konsep ini digunakan pada matriks simetrik yang terurai dan matriks diagonal yang didapatkan dengan mengurutkan dari kecil ke besar, seperti pada analisis komponen utama.

2.5.21. Bentuk Kuadratik

Misalkan **A** = (a_{ij}) adalah matriks berukuran n x n dan **x** adalah vektor peubah berukuran n x 1. Maka

$$\mathbf{x'Ax} = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij} x_i x_j$$

$$= a_{11}x_1^2 + \ldots + a_{nn}x_n^2 + (a_{12} + a_{21})\, x_1x_2 + \ldots + (a_{n-1,n} + a_{n,n-1})\, x_{n-1}x_n$$

Bentuk itu adalah polinomial derajat dua dari $x_1, \ldots, x_n$, sehingga disebut sebagai bentuk kuadratik dari **x.**

Misalkan,

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 2 & 1 \\ 1 & 1 & 1 \end{bmatrix} \text{ dan } \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

Maka

$$\mathbf{x'Ax} = x_1^2 + 2x_2^2 + x_3^2 + 6x_1x_2 + 4x_1x_3 + 2x_2x_3$$

Jelas bahwa **x'Ax** = **x'A'x**, dan dengan merata-ratakan juga sama dengan **x'**$\left(\frac{\mathbf{A}+\mathbf{A}'}{2}\right)$**x.** Karena $\left(\frac{\mathbf{A}+\mathbf{A}'}{2}\right)$ selalu simetrik, maka tanpa mengurangi maknanya bentuk kuadratik yang didefinisikan di atas ditambahkan bahwa **A** merupakan matriks simetrik. Untuk contoh sebelumnya, bentuk **x'Ax** dengan

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 2 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

dapat dituliskan **x'Ax = x'Bx** dengan

$$\mathbf{B} = \left(\frac{\mathbf{A}+\mathbf{A}'}{2}\right) = \begin{bmatrix} 1 & 3 & 2 \\ 3 & 2 & 1 \\ 2 & 1 & 1 \end{bmatrix}$$

Pada kasus **A** merupakan matriks simetrik, maka bentuk kuadratik dapat ditampilkan menggunakan salah satu dari pernyataan berikut

$$\mathbf{x'Ax} = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}x_i x_j$$

$$= \sum_{i=1}^{n} a_{ii}x_i^2 + \sum_{i=1}^{n}\sum_{j=1(i\neq j)}^{n} a_{ij}x_i x_j$$

$$= \sum_{i=1}^{n} a_{ii}x_i^2 + 2\sum_{i=1}^{n}\sum_{j=1(i<j)}^{n} a_{ij}x_i x_j$$

Persamaan

$$\mathbf{x'Ax} = c,$$

dengan c sebuah konstanta, menggambarkan permukaan kuadratik pada ruang berdimensi-n. Jadi dapat berupa parabola, hiperbola, atau elipsoida, tergantung pada unsur-unsur **A**. Bentuk elipsoida merupakan bentuk khusus yang sering dijumpai di statistika, dan ini terjadi ketika **B** adalah matriks definit positif atau semidefinit positif.

2.5.22. Matriks Definit dan Semidefinit Positif

Sebuah matriks simetrik berukuran $n \times n$, **A** dikatakan bersifat definit positif jika untuk sembarang vektor $\mathbf{x} \neq \mathbf{0}$, bentuk kuadratik $\mathbf{x'Ax} > 0$. Hampir mirip dengan itu, dikatakan semidefinit positif jika $\mathbf{x'Ax} \geq 0$. Jika **A** adalah definit

positif maka persamaan **x'Ax** = *c*, dengan *c* adalah konstanta, akan berupa elipsoida. Jika **A** adalah matriks diagonal yang semua unsur diagonalnya bernilai positif, maka **A** adalah matriks definit positif, tapi jika ada paling tidak sebuah unsur bernilai 0 (yang lain positif) menjadi matriks semi definit positif. Matriks diagonal yang unsurnya adalah ragam peubah, akan bersifat demikian karena ragam tidak pernah bernilai negatif.

Sudah diketahui bahwa untuk matriks yang definit positif, akar-akar cirinya semua bernilai positif. Demikian pula sebaliknya. Karena itulah maka determinan dari matriks definit positif juga bernilai positif, karena berupa hasil perkaliannya. Jadi determinannya tidak sama dengan nol, sehingga **A** bersifat non-singular.

Untuk matriks **B** berukuran *m x n*, maka **BB'** dan **B'B** bersifat semidefinit positif. Jika $m < n$ dan r(**B**) = m maka **BB'** definit positif, tapi **B'B** masih semidefinit positif.

2.5.23. Akar Kuadrat dari Matriks Simetrik Semi Definit Positif

Untuk sebuah matriks simetrik semidefinit positif **A**, dapat diperoleh matriks segitiga atas **U** sehingga

$$\mathbf{A} = \mathbf{U}'\mathbf{U}$$

Persamaan di atas disebut penguraian Cholesky. Matriks **U** merupakan matriks segitiga atas, sehingga tidak simetrik.

Akar kuadrat dari matriks simetrik, $\mathbf{A}^{1/2}$ dapat diperoleh melalui penguraian spektral

$$\mathbf{A} = \mathbf{P}\Lambda\mathbf{P}' = (\mathbf{P}\Lambda^{1/2}\mathbf{P}')\,(\mathbf{P}\Lambda^{1/2}\mathbf{P}') = \mathbf{A}^{1/2}\mathbf{A}^{1/2},$$

dengan **P** adalah matriks ortogonanl dan Λ matriks diagonal. Matriks diagonal Λ berisi akar ciri matriks **A** pada unsur diagonalnya, dan karena akar cirinya semuanya positif maka $\Lambda^{1/2}$ adalah matriks diagonal yang unsur diagonalnya akar kuadarat dari akar ciri. $\mathbf{A}^{1/2}$ diperoleh dari ($\mathbf{P}\Lambda^{1/2}\mathbf{P}'$) sehingga $\mathbf{A}^{1/2}$ bersifat simetrik.

2.5.24. Penguraian Nilai Singular (*Singular Value Decomposition*)

Sembarang matriks **B** berukuran *m x n* dapat dituliskan menjadi **B** = **UQV'**, dengan **U** dan **V** bersifat ortogonal atau sub ortogonal. Jika m lebih besar daripada *n*, maka **U** sub-ortogonal, dan **V** ortogonal. Matriks **Q** adalah matriks diagonal berukuran *n x n*. Jika *m* lebih kecil daripada *n*, maka setelah membuang (n-m) kolom nol pada **U** menjadi $\mathbf{U}_*$ dan **V** yang keduanya ortogonal. Jika **B** merupakan matriks persegi, maka baik **U** dan **V** ortogonal. Unsur diagonal matriks **Q** adalah nilai singular matriks **B**.

Penguraian nilai singular ini juga dapat dituliskan dalam bentuk kedua matriks di sebelah kanan dan kiri tidak hanya sub-ortogonal tapi sudah ortogonal. Pada kasus ini, matriks **Q** berukuran *m x n*. Jika *m* = *n*, tidak ada yang perlu dilakukan karena **U** dan **V** sudah ortogonal. Jika *m* > *n*, kita tuliskan **B** sebagai

$$\mathbf{B} = \left[\mathbf{U}_{mxn} \mid \mathbf{U}_{c_{mx(m-n)}}\right] \begin{bmatrix} \mathbf{Q}_{nxn} \\ --- \\ \mathbf{0}_{(m-n)xn} \end{bmatrix} \mathbf{V}'_{nxn}$$

$$= \mathbf{U}_* \mathbf{Q}_* \mathbf{V}_*'$$

Dalam hal ini $\mathbf{V}_* = \mathbf{V}$, $\mathbf{Q}_* = \begin{bmatrix} Q_{mxn} \\ --- \\ 0_{(m-n)xn} \end{bmatrix}$ dan $\mathbf{U}_* = [\mathbf{U} \mid \mathbf{U}_C]$. Matriks $\mathbf{U}_C$ dipilih sehingga $\mathbf{U}_C\mathbf{U} = \mathbf{0}$. Matriks $\mathbf{U}_C$ ini disebut sebagai matriks komplemen ortogonal dari matriks **U**.

Jika $m < n$, maka matriks **U** yang berukuran *m x n* memiliki (*n-m*) vektor nol. Matriks $\mathbf{U}_*$ diperoleh dengan membuang kolom-kolom tersebut. $\mathbf{V}_*$ sama dengan **V**, dan $\mathbf{Q}_*$ sama dengan **Q**.

2.5.25. Penguraian Nilai Singular Umum (*Generalized Singular Value Decomposition*)

Pada penguraian nilai singular matriks **B** di atas, matriks $\mathbf{U}_*$ dan $\mathbf{V}_*$ ortogonal, sehingga

$$\mathbf{U}_*^{'}\mathbf{U}_* = \mathbf{U}_*\mathbf{U}_*^{'} = \mathbf{I}_m$$

dan

$$\mathbf{V}_*^{'}\mathbf{V}_* = \mathbf{V}_*\mathbf{V}_*^{'} = \mathbf{I}_n$$

Jika membuang tanda *, kita ganti dengan

$$\mathbf{U'CU} = \mathbf{I}_m$$

dan

$$\mathbf{V'DV} = \mathbf{I}_n$$

dimana **C** dan **D** berturut-turut adalah matriks simetrik definit positif berukuran *m x m* dan *n x n*. Kita masih memiliki penguraian

$$\mathbf{B} = \mathbf{UQV'},$$

dengan **U** dan **V** memenuhi dua persyaratan tambahan. Penguraian ini dikenal sebagai penguraian nilai singular umum dari matriks **B**. Penguraian seperti ini akan berguna pada analisis korespondensi.

2.5.26. Perkalian Kronecker

Hasil perkalian Kronecker **C** dengan **D** (dinotasikan **C** ⊗ **D**) diperoleh dengan mengalikan setiap unsur matriks **C** dengan matriks **D**, dan kemudian membuat matriks gabungannya. Dalam notasi, perkalian Kronecker didefinisikan **C** ⊗ **D** = (c_{ij}**D**).

Misalkan

$$\mathbf{C} = \begin{bmatrix} 1 & 0 & 3 & 4 \\ 0 & 4 & 1 & -1 \\ 1 & 1 & -3 & 2 \end{bmatrix} \text{ dan } \mathbf{D} = \begin{bmatrix} 1 \\ 3 \\ 7 \end{bmatrix}$$

maka hasil perkalian Kronecker **C** ⊗ **D** adalah

$$\mathbf{C} \otimes \mathbf{D} = \begin{bmatrix} 1 & 0 & 3 & 4 \\ 3 & 0 & 9 & 12 \\ 7 & 0 & 21 & 28 \\ 0 & 4 & 1 & -1 \\ 0 & 12 & 3 & -3 \\ 0 & 28 & 7 & -7 \\ 1 & 1 & -3 & 2 \\ 3 & 3 & -9 & 6 \\ 7 & 7 & -21 & 14 \end{bmatrix}$$

2.6. Latihan

1. Terdapat tiga buah vector **x, y z** yang didefinisikan sebagai berikut:

   x' = [1, -1, 2]

   y' = [2, 0, -1]

   z' = [0, -2, 5]

   a. Apakah vektor-vektor tersebut di atas bebas linier, buktikan jawaban anda!
   b. Bila tidak bebas linier, buatlah vektor tertentu yang merupakan kombinasi linier dari vektor yang lain.

2. Dua buah vektor, katakanlah

   x' = [-1, 5, 2, -2]

   y' = [4, -3, 0, 1]

   Berapa sudut yang dibentuk oleh dua vektor tersebut?

3. Suatu bentuk kuadrat

   $$2X_{12} + 2X_{22} + 2X_{32}+2X_1X_2+2X_1X_3+2X_2X_3$$

   a. Tentukan matriks bentuk kuadrat di atas
   b. Tentukan akar ciri dari matriks nomor a tersebut
   c. Apakah matriks tersebut matriks definit positip

4. Suhu suatu tempat (katakan disimbolkan sebagai Y) ditentukan oleh dua variabel utama, katakan X_1 dan X_2. Tiga data tentang ketinggian dan nilai X_1 dan X_2 adalah sebagai berikut :

X_1	X_2	Y
4	12	44
5	14	52
7	16	62

Seorang mahasiswa ingin membuat suatu fungsi yang menghubungkan antara Y dengan X_1 dan X_2, katakan persamaan tersebut adalah

$$Y = a X_1 + bX_2$$

Tentukan nilai a dan b persamaan diatas dengan menggunakan sistem Persamaan Linier.

3

3.Sebaran Normal Ganda (*Multivariate Normal Distribution*)

3.1. Peubah Acak

Dari suatu percobaan seringkali kita lebih tertarik pada suatu fungsi hasil percobaannya, bukan pada hasil percobaannya itu sendiri. Misalnya dalam pelemparan dua dadu kita sering tertarik pada jumlah mata dadu yang muncul dari kedua dadu, bukan hasil dari masing-masing dadu tersebut. Dengan kata lain, kita mungkin tertarik ingin tahu apakah jumlahnya 5 dan tidak peduli apakah hasil percobaan yang sesungguhnya dadu yang muncul (1,4) atau (2,3) atau (3,2) atau (4,1). Besaran-besaran yang menjadi perhatian kita ini, atau lebih formalnya , fungsi bernilai real yang didefinisikan pada ruang contoh percobaan ini, di kenal sebagai peubah acak (random variabel). Atau dengan kata lain peubah acak adalah suatu fungsi yang memetakan unsur-unsur ruang contoh suatu percobaan terhadap suatu gugus bilangan-bilangan nyata sebagai wilayah fungsi.

3.2. Fungsi Sebaran

Yang di maksud dengan fungsi sebaran atau fungsi sebaran kumulatif suatu peubah acak **X** ialah:

$$F(x) = P(X \leq x)$$

untuk semua bilangan nyata x, - ∞ < x < ∞ . Dengan kata lain, F(x) menyatakan peluang bahwa peubah acak **X** mengambil nilai lebih kecil atau sama dengan x. Beberapa sifat fungsi sebaran kumulatif adalah :

1. F adalah suatu fungsi tidak turun (nondecreasing function); artinya jika a < x , maka F (a) ≤ F (x).
2. $\lim_{x \to \infty} F(x) = 1$
3. $\lim_{x \to -\infty} F(x) = 0$
4. F (x) kontinu kanan, artinya $\lim_{x \to x_0^+} F(x) = F(x_0)$

3.3. Peubah Acak Diskret

Peubah acak diskret adalah peubah acak yang himpunan semua kemungkinan nilai yang dapat diambilnya terhingga atau tak hingga tercacah. Bagi suatu peubah acak diskret X, kita definisikan fungsi massa peluang P(a) sebagai :

P(a) = P (X = a)

Fungsi massa peluang P(a) bernilai positif untuk paling banyak sejumlah tercacah nilai a. Dengan kata lain, jika X mengambil salah satu dari nilai-nilai x1, x2, ... maka:

P (xi) ≥ 0i = 1, 2, ...

P (x) = 0 semua nilai x lainnya

Karena X pasti mengambil salah satu dari nilai xi , maka :

$$\sum_{i=1}^{\sim} p(x_i) = 1$$

Contoh-contoh peubah acak diskret antara lain peubah acak Bernoulli, peubah acak binom, peubah acak Poisson dan lain-lain.

3.4. Peubah Acak Kontinu

Peubah acak kontinu adalah peubah acak yang himpunan semua kemungkinan nilai yang dapat diambilnya tak tercacah. Misalkan **X** adalah peubah acak kita katakan bahwa **X** adalah suatu peubah acak kontinu jika ada suatu fungsi tak negatif f, yang terdefinisikan untuk semua bilangan nyata x ∈ (- ∞ , ∞), dengan sifat bahwa untuk sembarang himpunan bilangan nyata R,

$$P\{X \in R) = \int_R f(x)dx$$

Fungsi f ini dinamakan fungsi kepekatan peluang bagi peubah acak **X**. Dengan kata lain persamaan di atas mengatakan bahwa peluang X ada di dalam R dapat diperoleh dengan cara mengintegralkan fungsi kepekatan peluangnya pada himpunan R.

Karena x pasti mengambil suatu nilai, maka f pasti memenuhi :

$$P\{\overline{\underline{X}} \in B\} = \int_{-\sim}^{\sim} f(x)dx = 1$$

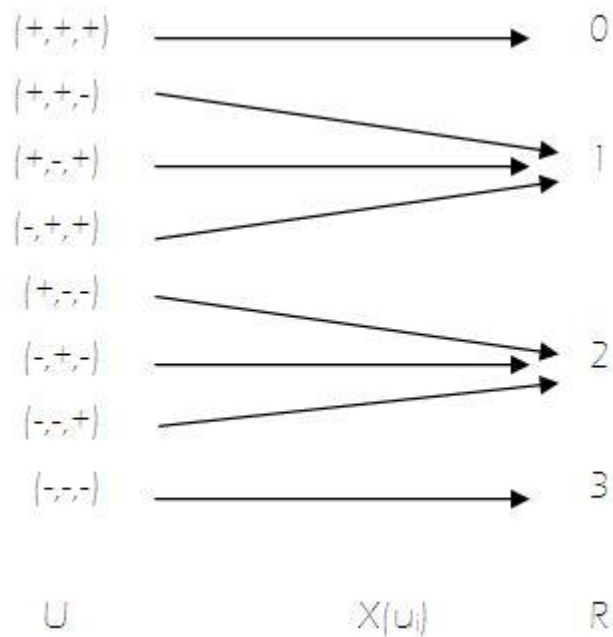
Contoh-contoh peubah acak kontinu antara lain peubah acak normal, seragam, weibull, gamma, dan lain-lain.

3.5. Peubah Acak Ganda

Peubah acak adalah suatu fungsi yang memetakan unsur-unsur ruang contoh suatu percobaan terhadap suatu gugus bilangan–bilangan nyata sebagai wilayah fungsi. Misalkan peubah acak tersebut dilambangkan sebagai X, dengan U={u_1, u_2, ..., u_n } sebagai daerahnya, maka yang dimaksudnya sebagai X(u_i) ialah suatu unsur yang merupakan bayangan

unsur $u_i \in U$, semua unsur $X(u_i)$ ini terkandung di dalam wilayah peubah acak X, yaitu $W_x \subset R$. Misalkan 3 keping mata uang dilempar dengan serentak, maka sebagai hasilnya akan termasuk dalam ruang contoh U,

U = {(+,+,+),(+,-,-),(+,-,+),(-,+,+),(+,-,-),(-,+,-),(-,-,+),(-,-,-)}

Apabila sisi – yang diperhatikan, maka unsur (+,+,+) di dalam u diberi nilai 0, karena tidak satu pun sisi – yang muncul. Sehingga dapat kita buat pemetaan sebagai berikut :

U adalah daerah pemetaan dan R adalah wilayahnya. Gugus R juga dapat dipandang sebagai ruang contoh percobaan kalau hasil suatu percobaan tidak dibatasi sebagai perincian letak ketiga macam uang yang dilempar, tetapi sebagai jumlah sisi yang – yang muncul. Dengan

memberi nilai seperti ini, ruang contoh baru R ini terdiri dari bilangan cacah yang merupakan bagian dari bilangan nyata R. Pemetaan dari U ke R merupakan suatu contoh mengenai peubah acak.

Namun seringkali hasil suatu percobaan tidak cukup dinyatakan dengan hanya satu macam sifat. Dengan mengetahui beberapa ciri suatu benda atau hasil dari suatu percobaan, kita berharap dapat mengetahui lebih banyak sifat benda atau hasil percobaan tersebut. Dalam hal ini kita menghadapi sifat ganda suatu benda yang seringkali berkaitan satu sama lain. Peubah acak tersebut adalah peubah acak ganda-p (peubah ganda-p, vektor peubah acak berdimensi p).

3.6. Pengertian Peubah Acak Ganda

Misalkan $X_1, X_2, \ldots, X_p$ merupakan p buah peubah acak yang didefinisikan pada ruang ukuran peluang (U,β,p) yang sama. Setiap peubah acak X_i, i=1,2, ... p memetakan U ke R.

Vektor $\mathbf{X}' = (X_1,X_2, \ldots, X_p)$ yang memetakan u ke R^p dimana $\mathbf{X}'(u) = (X_1(u),\ldots, X_p(u))$, $u \in U$ disebut peubah acak ganda-p, jika bayangan kebalikan setiap selang berdimensi-p :

$$I = \{(X_1,X_2, \ldots, X_p) ; -\infty < X_i < r_i,\ r_i \in R,\ i = 1,2, \ldots p \}$$

merupakan kejadian pada β, yaitu jika :

$$\mathbf{X}^{-1}(u)= \{ u ; X_1(u) < r_1, \ldots, X_p(u) < r_p \} \in \beta \text{ untuk setiap } r_i \in R.$$

Misalkan peubah acak **X** berdimensi p didefinisikan sebagai vektor :

$$\mathbf{X}' = [X_1,X_2, \ldots, X_p]$$

Notasi ragam untuk vektor acak p komponen adalah matriks ragam peragam :

$$Cov(X, X') = E\{[X - E(X)][X - E(x)]'\}$$

$$= \begin{bmatrix} \sigma_{11} & \dots & \sigma_{1p} \\ \dots & \dots & \dots \\ \sigma_{1p} & \dots & \sigma_{pp} \end{bmatrix}$$

$$= \Sigma$$

Kita sering menamai matriks simetrik tersebut dengan matriks kovarian dari **X**. Sedangkan ragam untuk kombinasi linear **a'X** = $a_1X_1 + \dots + a_pX_p$ dari peubah acak adalah ;

$$Var(a'X) = \sum_{i=1}^{p}\sum_{j=1}^{p} a_i a_j \sigma_{ij}$$

$$= \boldsymbol{a}'\Sigma\boldsymbol{a}$$

Kovarian dari dua kombinasi linear **a'X** dan **b'X** adalah :

$$Cov(a'X, b'X) = \sum_{i=1}^{p}\sum_{j=1}^{p} a_i b_j \sigma_{ij}$$

$$= \boldsymbol{a}'\Sigma\boldsymbol{b}$$

Secara umum, jika A dan B mempunyai dimensi r x p dan s x p maka kovarian dari transformasi : Y = AX dan Z = BX adalah matriks dimana :

$$Cov(Y, Y') = A\Sigma A'$$

$$Cov(Z, Z') = B\Sigma B'$$

$$Cov(Y, Z') = A\Sigma B'$$

R = matriks korelasi dari populasi :

$$R = \begin{bmatrix} 1 & \rho_{12} & \dots & \rho_{1p} \\ \rho_{12} & 1 & \dots & \rho_{2p} \\ \dots & \dots & \dots & \dots \\ \rho_{1p} & \rho_{2p} & \dots & 1 \end{bmatrix}$$

dimana $D(\sigma_i)$ adalah matriks diagonal yang unsur-unsurnya merupakan simpangan baku dari peubah acak ke-i, maka matriks kovarian dan korelasi mempunyai hubungan sebagai berikut :

$$R = D\left(\frac{1}{\sigma_i}\right) \Sigma D\left(\frac{1}{\sigma_i}\right)$$

$$\Sigma = D(\sigma_i) R D(\sigma_i)$$

3.7. Sebaran Normal Ganda Dan Sifat-Sifatnya

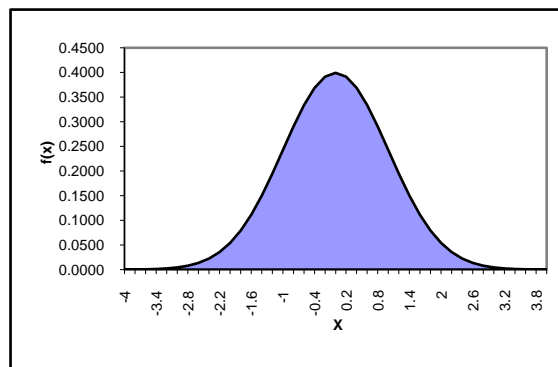
Fungsi kepekatan normal ganda (multivariate normal) adalah generalisasi dari fungsi kepekatan univariate normal dengan $p \geq 2$ dimensi. Fungsi kepekatan dari peubah acak x yang menyebar normal dengan nilai tengah μ dan ragam σ^2 adalah:

$$f(x) = \frac{1}{\sqrt{2\Pi\sigma^2}} e - [(x - \mu)/\sigma]^2 / 2$$

dimana $-\infty < x < \infty$.

Plot dari fungsi ini akan menghasilkan kurva berbentuk genta (Gambar 3.1) yang memiliki ciri-ciri sebagai berikut:

1. Simetrik terhadap nilai tengah (μ)
2. Mean, median, modus berada pada titik yang sama
3. P (μ - θ < x < μ + σ) = 0.683
4. P (μ - 2θ < x < μ + 2σ) = 0.954

Fungsi kepekatan normal dengan nilai tengah μ dan ragam σ^2 dinotasikan sebagai N (μ,σ^2), notasi ini akan digunakan pada kasus multivariate yang akan dibahas selanjutnya.

Gambar 3.1 Sebaran normal dengan nilai tengah μ dan ragam σ^2

Fungsi kepekatan bersama dari peubah acak yang menyebar normal dan saling bebas adalah:

$$f(x_1,\ldots,x_p) = \frac{1}{(2\pi)^{p/2}\sigma_1 .. \sigma_p} Exp\left[-1/2\sum_{i=1}^{p}\left(\frac{x_i-\mu_i}{\sigma_i}\right)\right]^2$$

bentuk [(x_i - μ)/σ]^2 dari eksponen fungsi sebaran normal mengukur jarak kuadrat dari x_i ke μ dalam unit simpangan baku. Bentuk ini dapat

digeneralisasikan untuk px1 vektor **x** dari pengamatan beberapa peubah sebagai:

$$(\mathbf{x} - \boldsymbol{\mu})' \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu})$$

dimana:

x' = [x1, ... , xp]   vektor peubah

$\mu' = [\mu_1, \ldots, \mu_p]$ *nilai tengah dari vektor acak x*

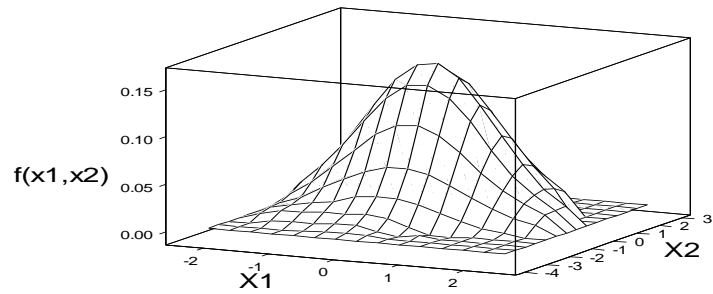
Σ = matriks ragam peragam (*covariance*)

$$\Sigma = \begin{bmatrix} \sigma_1^{\ 2} \ldots 0 \\ \\ 0 .. \sigma_p^{\ 2} \end{bmatrix}$$

maka fungsi kepekatan peluang bersamanya dapat ditulis sebagai berikut :

$$f(\underline{x}) = \frac{1}{(2\pi)^{p/2} |\Sigma|^{1/2}} Exp \left[-1/2(\underline{x} - \underline{\mu})' \Sigma^{-1} (\underline{x} - \underline{\mu}) \right]$$

dimana $-\infty < x_i < \infty$, i = 1,2,...p. Fungsi kepekatan normal berdimensi p ini dapat ditulis sebagai $N_p(\boldsymbol{\mu},\Sigma)$ yang analog dengan kasus univariate.

Sebagai ilustrasi disajikan gambar fungsi kepekatan normal ganda 2 dengan pusat di $\mu_1=\mu_2=0$, $\sigma_{12}=\sigma_{22}=1$ dan $\sigma_{12}=0$.

Gambar 3.2 Sebaran normal ganda 2 dengan μ_1=μ_2=0, σ_{12}=σ_{22}=1 dan σ_{12}=0

3.8. Kontur

Eksponen (x - μ)' Σ-1 (x - μ) dari kepekatan normal ganda memperlihatkan persamaan elipsoid dalam ruang peubah berdimensi p jika berbantuk ini, ditulis dalam sebuah persamaan terhadap sebuah nilai konstanta positif c. Keluarga elipsoid ini dibangkitkan dari konstanta c yang bervariasi dengan nilai tengah μ.

Kontur dari fungsi peluang kepekatan untuk nilai konstanta

c = {semua x yang memenuhi (x - μ)'Σ-1 (x - μ) = c2}

= permukaan ellips berpusat pada μ

absis bagi setiap ellip dari kepekatan yang konstan adalah arah vektor ciri dari Σ-1 dan panjangnya porporsional terhadap akar kuadrat dari nilai akar ciri Σ-1 .

Teorema:

Jika Σ definisi positif maka Σ-1 ada, Σ e = λ e berimplikasi Σ-1 e = (1/λ)e, sehinggga (λ.e) adalah pasangan nilai akar ciri dan vektor ciri bagi Σ koresponden terhadap pasangan (1/λ,e) untuk Σ-1 . Σ -1 juga positif.

Kontur dari kepekatan yang konstan dari sebaran normal berdimensi p adalah ellipsoid, sedemikian sehingga:

$$(x - \mu)' \Sigma\text{-}1\ (x - \mu) = c2$$

dengan pusat ellips adalah μ dan absis $\pm$ c $\sqrt{}$ λi ei , dimana

$$\Sigma\ ei = \lambda i\ ei \ ,\ i = 1,2,...p.$$

Sifat-sifat sebaran normal ganda :

- Kombinasi linier dari semua komponen peubah x juga menyebar normal. Jika $X_p \sim N_p$ ($\underline{\mu}$, Σ) , maka kombinasi linear dari peubah a'X = $a_1 X_1 + a_2 X_2 + ... + a_p X_p$ menyebar N (a' μ , a' Σ a)

- Jika $X_p \sim N_p$ (μ , Σ) maka semua anak gugus dari X juga menyebar normal. Semua anak gugus X menyebar normal. Jika X dipartisi, vektor nilai tengah μ , dan kovarian matriks Σ , sebagai berikut:

$$\underset{(p\times 1)}{X} = \begin{bmatrix} \underset{(a\times 1)}{x_1} \\ \hline \underset{((p-a)x_1)}{x_2} \end{bmatrix}, \quad \mu = \begin{bmatrix} \underset{(q\times 1)}{\mu_1} \\ \hline \underset{((p-q)x_1)}{\mu_2} \end{bmatrix}, \quad \underset{(pxp)}{\sum} = \begin{bmatrix} \underset{(qxq)}{\Sigma_{11}} & \underset{(qx(p-q))}{\Sigma_{12}} \\ \hline \underset{((p-q)xq)}{\Sigma_{21}} & \underset{((p-q)x(p-q))}{\Sigma_{22}} \end{bmatrix}$$

dan

$$N_{q1+q2}\left(\begin{bmatrix} \mu_1 \\ \hline \mu_2 \end{bmatrix}, \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \hline \Sigma_{21} & \Sigma_{22} \end{bmatrix} \right)$$

maka X1 ~ Nq (μ1,σ11).

- Kovarian bernilai nol mengimplikasikan komponen yang berpadanan saling bebas.

  - Jika X_1 dan X_2 saling bebas, maka kovarian (X_1, X_2) = 0

  - Jika X_1 dan X_2 saling bebas, dan menyebar N_{q1} (μ_1, Σ_{11}) dan N_{q2} (μ_2, Σ_{22}) maka sebaran bersyarat [x_1 | x_2] adalah normal ganda :

$$N_{q1+q2}\left(\begin{bmatrix}\mu_1\\ \mu_2\end{bmatrix},\begin{bmatrix}\Sigma_{11} & 0\\ 0 & \Sigma_{22}\end{bmatrix}\right)$$

- Sebaran bersyarat dari semua peubah menyebar normal ganda: $X=\begin{bmatrix}x_1\\ x_2\end{bmatrix}\sim N_p(\underline{\mu},\Sigma)$ dengan $\mu=\begin{bmatrix}\mu_1\\ \mu_2\end{bmatrix},\Sigma\begin{bmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{21} & \Sigma_{22}\end{bmatrix}$ dan | Σ_{22} | > 0. Maka sebaran bersyarat X dengan X_2 = x_2 adalah normal dengan nilai tengah = $\mu_1 + \Sigma_{12}\,\Sigma_{22}^{-1}\,(x_2 - \mu_2)$ dan kovarian $\Sigma_{11} - \Sigma_{12}\,\Sigma_{22}^{-1}\,\Sigma_{21}$.

- Dua sifat terakhir dari sebaran normal ganda. Sebaran X^2 menentukan keragaman dari ragam contoh $S^2 = S_{11}$ untuk contoh dari sebaran normal. Hal ini juga penting dalam kasus normal ganda. Jika X ~ N_p (μ,Σ) dengan | Σ | > 0 maka:

  a. (x - μ) Σ^{-1} (x - μ) ~ X_p^2 dimana X_p^2 menyatakan sebaran khi kuadrat dengan derajat bebas p.

  b. Selang kepercayaan (1 - α) $\rightarrow$ (x - μ)' Σ^{-1} (x - μ) = $\chi^2_{(\alpha,p)}$.

3.9. Eksplorasi Sebaran Normal Ganda

Untuk mengevaluasi apakah gugus data yang dimiliki menyebar normal ganda dapat ditelusuri dengan cara ekplorasi. Seperti halnya untuk kasus univariate penelusuran sebaran normal ganda dapat juga memanfaatkan plot quantil-quantil. Plot quantil-quantil yang digunakan dalam kasus univariate adalah quantil normal sedangkan dalam kasus multivariate plot quantil-quantil didekati dengan quantil khi-kuadrat.

Beberapa tahapan yang harus dilakukan dalam menyusun Plot Kuartil χ^2 adalah sebagai berikut:

1. Hitung:

$$d_{ii}^{\ 2} = (\underline{x}_{(i)} - \mu)' \Sigma^{-1} (\underline{x}_{(i)} - \mu)$$

2. Beri peringkat nilai d_{ii}^2
3. Carilah nilai khi-kuadrat dari nilai (i –1/2)/n dengan derajat bebas p.

$$\chi_p^2\left(\frac{i - \frac{1}{2}}{n}\right)$$

4. Buat plot $\chi_p^2\left(\frac{i - \frac{1}{2}}{n}\right)$ dengan d_{ii}^2, bila pola hubungannya mengikuti garis lurus maka data tersebut dapat dikatakan menyebar normal ganda. Namun demikian untuk lebih menyakinkan dapat dilakukan dengan menghitung nilai korelasi person

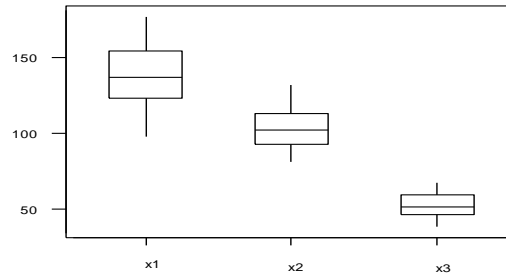
$$\chi_p^2\left(\frac{i - \frac{1}{2}}{n}\right)$$

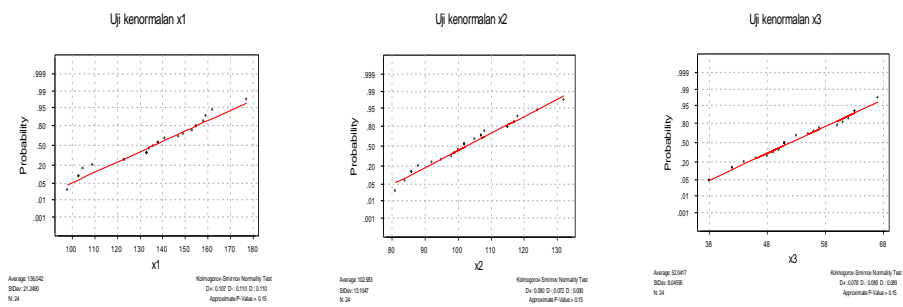
dengan d_{ii}^2. Apabila nilai korelasi ini nyata maka data tersebut mengikuti sebaran normal ganda.

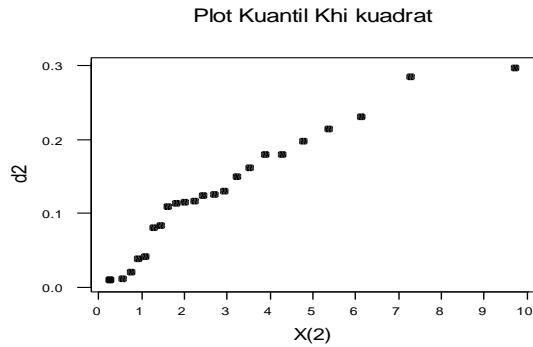
Contoh:

Dari suatu pengamatan diperoleh data sebagai berikut:

No.	X_1	X_2	X_3	No.	X_1	X_2	X_3
1	98	81	38	13	138	98	51
2	103	84	38	14	138	99	51
3	103	86	42	15	141	105	53
4	105	86	42	16	147	108	57
5	109	88	44	17	149	107	55
6	123	92	50	18	153	107	56
7	123	95	46	19	155	115	63
8	133	99	51	20	155	117	60
9	133	102	51	21	158	115	62
10	133	102	51	22	159	118	63
11	134	100	48	23	162	124	61
12	136	102	49	24	177	132	67

Secara eksplorasi ketiga peubah yang diamati tidak ada yang aneh, bahkan dari diagram kotak garis terlihat bahwa ketiga peubah menyebar simetrik (Gambar 3.3).

Gambar 3.3 Diagram kotak garis untuk peubah X1, X2 dan X3

Sementara pada Gambar 3.4 disajikan plot kuantil ketiga peubah membentuk garis lurus yang mengindikasikan ketiga peubah menyebar normal, hal ini juga didukung oleh hasil pengujian kenormalan Kolmogorov-Smirnov dimana nilai p_value yang lebih besar dari 5%. Jadi untuk masing-masing peubah tersebut menyebar Normal.

Gambar 3.4 Uji kenormalan Kolmogorov-Smirnov untuk peubah X1, X2 dan X3

Plot kuantil-kuantil khi-kuadrat dari ketiga peubah menunjukkan pola yang linier sehingga dapat dapat disimpulkan ketiga peubah tersebut menyebar Normal Ganda. Perhitungan d^2 dan $\chi^2(p)$ secara lengkap dapat dilihat pada Tabel 3.1.

Gambar 3.5 Plot quantil khi-kuadrat untuk peubah X1, X2 dan X3

Untuk membuktikan apakah plot tersebut benar-benar menunjukkan Normal Ganda maka dicari korelasi antara d^2 dengan $\chi^2(p)$ yaitu r_Q = 0.962442 yang lebih besar dari batas kritis pada taraf nyata 5% yaitu 0.9564. Ini menunjukkan bahwa d^2 dan $\chi^2(p)$ memiliki hubungan linier yang nyata.

Tabel 3.1 Perhitungan d^2 dan $\chi^2(p)$

Amatan	x1	x2	x3	d^2	Peringkat d^2	P= (i-1/2)/n	$\chi^2_{(p)}$
1	98	81	38	0.149	16	0.646	3.254
2	103	84	38	0.179	19	0.771	4.317
3	103	86	42	0.179	18	0.729	3.915
4	105	86	42	0.123	13	0.521	2.479
5	109	88	44	0.114	11	0.438	2.048
6	123	92	50	0.213	21	0.854	5.383
7	123	95	46	0.040	6	0.229	1.126
8	133	99	51	0.019	4	0.146	0.780
9	133	102	51	0.009	1.5	0.042	0.309

Amatan	x1	x2	x3	d^2	Peringkat d^2	P= (i-1/2)/n	$\chi^2_{(p)}$
10	133	102	51	0.009	1.5	0.042	0.309
11	134	100	48	0.128	15	0.604	2.973
12	136	102	49	0.115	12	0.479	2.257
13	138	98	51	0.160	17	0.688	3.564
14	138	99	51	0.113	10	0.396	1.850
15	141	105	53	0.011	3	0.104	0.603
16	147	108	57	0.038	5	0.188	0.954
17	149	107	55	0.083	8	0.313	1.477
18	153	107	56	0.196	20	0.813	4.794
19	155	115	63	0.230	22	0.896	6.158
20	155	117	60	0.079	7	0.271	1.300
21	158	115	62	0.108	9	0.354	1.660
22	159	118	63	0.125	14	0.563	2.716
23	162	124	61	0.283	23	0.938	7.315
24	177	132	67	0.296	24	0.979	9.748

3.10. Pengambilan Contoh Dari Sebaran Normal Ganda

3.10.1. Likelihood Normal Ganda

Diasumsikan p x 1 vektor $X_1, X_2, X_3, \ldots, X_n$ adalah contoh acak dari populasi normal ganda dengan vektor rataan μ dan matrik covariance Σ, $X_1, X_2, X_3, \ldots, X_n$ saling bebas dan masing-masing menyebar $N_p(\mu,\Sigma)$, fungsi kepekatan bersama untuk seluruh observasi ini adalah perkalian dan kepekatan normal ganda.

Jadi,

$$\begin{Bmatrix} kepeka\tan bersama \\ dariX_1, X_2, ..., X_n \end{Bmatrix} = \prod_{j=1}^{n} \left\{ \frac{1}{(2\pi)^{p/2} |\Sigma|^{1/2}} e^{-(X_j-\mu)'\Sigma^{-1}(X_j-\mu)/2} \right\}$$

$$= \frac{1}{(2\pi)^{np/2}} \frac{1}{|\Sigma|^{n/2}} e^{-\sum_{j=1}^{n}(X_j-\mu)'\Sigma^{-1}(X_j-\mu)/2}$$

Ketika nilai numerik dari observasi tersedia, ini dapat disubsitusikan untuk X_j dalam persamaan di atas. Hasilnya, dipikirkan sebagai fungsi dari μ dan Σ untuk deret tetap observasi $X_1, X_2, X_3, \dots, X_n$ disebut sebagai **Likelihood.**

Banyak prosedur statistik yang menggunakan nilai data observasi terbaik yang menerangkan parameter populasi. Terbaik artinya memilih nilai parameter yang dapat memaksimumkan kepekatan bersama yang dievalusasi darti observasi data. Teknik ini disebut pendugaan maksimum likelihood dan nilai parameter yang dimaksimumkan disebut penduga maksimum likelihood.

Kita akan memikirkan penduga maksimum likelihood dari parameter μ dan Σ untuk populasi normal ganda, jadi, kita ambil contoh, $X_1, X_2, X_3, \dots, X_n$ yang tetap dan memikirkan kepekatan bersamanya dengan persamaan di atas.

Hasilnya berupa sebuah fungsi likelihood. Kita akan memerlukan beberapa tambahan sifat untuk teras matriks segi. (Teras suatu matriks adalah jumlah elemen diagonalnya).

Umpamakan A adalah k x k matriks simetrik dan **x** adalah vektor k x 1.

a. $\mathbf{x}'A\mathbf{x} = tr(\mathbf{x}'A\mathbf{x}) = tr(A\mathbf{x}\mathbf{x}')$

b. tr (A) = $\sum_{i=1}^{k} \lambda_i$, dimana λi adalah nilai eigen dari A

Bukti

Untuk bagian a, kita perhatikan bahwa **x**'A**x** adalah skalar, jadi **x**'A**x** = tr (**x**'A**x**), terlihat bahwa tr (BC) = tr(CB) untuk setiap 2 matrik B dan C dengan dimensi m x k dan k x m. Hal ini karena BC mempunyai $\sum_{j=1}^{k} b_{ij} c_{ji}$ sebagai elemen diagonal ke-I, sehingga

$$\text{tr (BC)} = \sum_{m=1}^{m} \left(\sum_{j=1}^{k} b_{ij} c_{ji} \right)$$

sama untuk elemen diagonal CB adalah $\sum_{i=1}^{m} c_{ji} b_{ij}$.

Jadi,

$$\text{tr (CB)} = \sum_{j=1}^{k} \left(\sum_{i=1}^{m} c_{ji} b_{ij} \right) = \sum_{i=1}^{m} \left(\sum_{j=1}^{k} b_{ij} c_{ji} \right) = \text{tr (BC)}.$$

Misal **X**' adalah matrik **B** dengan m=1 dan misal **Ax** memainkan aturan matrik **C**. Maka tr (**x**'(A**x**)) = tr ((A**x**)**x**') = tr (A**x x**').

Bagian b dibuktikan dengan dekomposisi spektral, untuk menulis A = P'AP, dimana PP' = I dan ∧ adalah matrik diagonal dengan masukan $\lambda_1, \lambda_2, \ldots, \lambda_k$. Oleh sebab itu, tr (A) = tr (P'∧P) = tr (∧PP') = tr (∧) = $\lambda_1 + \lambda_2 + \ldots + \lambda_k$.

Sekarang kepekatan bersama eksponen dalam persamaan dapat disederhanakan, yaitu :

$$(X_j-\mu)'\Sigma^{-1}(X_j-\mu)=tr\left[(X_j-\mu)'\Sigma^{-1}(X_j-\mu)\right]$$
$$=tr\left[\Sigma^{-1}(X_j-\mu)(X_j-\mu)'\right]$$

$$\sum_{j=1}^{n}(X_j-\mu)'\Sigma^{-1}(X_j-\mu)=\sum_{j=1}^{n}tr\left[(X_j-\mu)'\Sigma^{-1}(X_j-\mu)\right]$$

$$=\sum_{j=1}^{n}tr\left[\Sigma^{-1}(X_j-\mu)(X_j-\mu)'\right]$$

$$=tr\left[\Sigma^{-1}\left(\sum_{j=1}^{n}(X_j-\mu)(X_j-\mu)'\right)\right]$$

Karena teras dari jumlah suatu matrik adalah sama dengan jumlah teras matriks-matriks tersebut kita dapat menambahkan dan mengurangkan $\bar{X}=(1/n)\sum_{j=1}^{n}X_j$ pada tiap-tiap suku (X_j-μ) dalam $\sum_{j=1}^{n}(Xj-\mu)(Xj-\mu)'$ sehingga :

$$\sum_{j=1}^{n}(X_j-\bar{X}+\bar{X}-\mu)(X_j-\bar{X}+\bar{X}-\mu)'=\sum_{j=1}^{n}(Xj-\bar{X})(Xj-\bar{X})'+\sum_{j=1}^{n}(\bar{X}-\mu)(\bar{X}-\mu)'$$

$$=\sum_{j=1}^{n}(Xj-\bar{X})(Xj-\bar{X})'+n(\bar{X}-\mu)(\bar{X}-\mu)'$$

karena perkalian tiap suku, $\sum_{j=1}^{n}(Xj-\bar{X})(\bar{X}-\mu)'$ dan $\sum_{j=1}^{n}(\bar{X}-\mu)(Xj-\bar{X})'$, keduanya matriks nol. Kepekatan bersama contoh acak dari populasi normal ganda dapat dituliskan sebagai berikut :

$$\left\{ \begin{matrix} kepekatan\,bersama \\ dari X_1, X_2, ..., X_n \end{matrix} \right\} = (2\pi)^{-np/2} |\Sigma|^{-n/2} \exp\left\{ -tr\left[\Sigma^{-1} \left(\sum_{j=1}^{n} (X_j - \bar{X})(X_j - \bar{X})' + n(\bar{X} - \mu)(\bar{X} - \mu)' \right) \right] / 2 \right\}$$

Substitusi nilai-nilai X_1, X_2, X_3, … , X_n pada kepekatan bersama yang menghasilkan fungsi likelihood. Kita akan menunjukkan fungsi ini oleh $L(\mu,\Sigma)$ untuk menekankan fakta bahwa ini adalah fungsi dari (tidak diketahui) parameter populasi μ dan Σ. Jadi, ketika vektor X_j mengandung bilangan yang spesifik dari observasi yang sebenarnya, kita punya :

$$L(\mu,\Sigma) = \frac{1}{(2\pi)^{np/2} |\Sigma|^{n/2}} e^{-tr\left[\Sigma^{-1} \left(\sum_{j=1}^{n} \left(X_j - \bar{X} \right)\left(X_j - \bar{X} \right)' + n\left(\bar{X} - \mu \right)\left(\bar{X} - \mu \right)' \right) \right] / 2}$$

Akan baik sekali bila diakhir bagian ini untuk menuliskan bentuk eksponen dari fungsi likelihood dalam cara yang berbeda. Terutama, kita akan menggunakan identitas :

$$tr\left(\Sigma^{-1} \left(\sum_{j=1}^{n} \left(X_j - \bar{X} \right)\left(X_j - \bar{X} \right)' + n\left(\bar{X} - \mu \right)\left(\bar{X} - \mu \right)' \right)\right)$$

$$= tr\left[\Sigma^{-1} \left(\sum_{j=1}^{n} (X_j - \bar{X})(X_j - \bar{X})' + n(\bar{X} - \mu)(\bar{X} - \mu)' \right) \right]$$

$$= tr\left[\Sigma^{-1} \left(\sum_{j=1}^{n} (X_j - \bar{X})(X_j - \bar{X})' \right) \right] + n(\bar{X} - \mu)' \Sigma^{-1} (\bar{X} - \mu)$$

3.10.2. Pendugaan Maksimum Likelihood untuk μ dan Σ

Misal **B** adalah matriks simetri definit positif berukuran p x p dan skalar b > 0.

$$\frac{1}{|\Sigma|^b} e^{-tr(\Sigma^{-1}B)/2} \leq \frac{1}{|B|^b} (2b)^{pb} e^{-bp}$$

untuk seluruh $\sum_{pxp}$ yang definit positif, dengan persamaan yang ada hanya untuk

$$\Sigma = (1/2\ b)B$$

Bukti

Anggap $B^{1/2}$ adalah akan kuadrat dari B jadi $B^{1/2}\ B^{1/2} = B$, $B^{1/2} = I$ dan $B^{-1/2}\ B^{-1/2} = B^{-1}$, maka tr $(\Sigma^{-1}B)$ = tr $[(\Sigma^{-1}\ B^{1/2})\ B^{1/2}]$ = tr $[B^{1/2}(\Sigma^{-1}\ B^{1/2})]$. Anggap η adalah nilai eigen dan $B^{1/2}\ \Sigma^{-1}\ B^{1/2}$. Matriks ini positif definit karena $y'\ B^{1/2}\ \Sigma^{-1}\ B^{1/2}y = (B^{1/2y})'\ \Sigma^{-1}(B^{1/2}y) > 0$ jika $B^{1/2}y \neq 0$ atau sama juga, $y \neq 0$. Jadi nilai eigen η_I dan $B^{1/2}\Sigma^{-1}B^{1/2}$ adalah positif. Memberikan hasil : tr $(\Sigma^{-1}B)$ = tr $(B^{1/2}\Sigma^{-1}B^{1/2}) = \sum_{i=1}^{p} \eta_i$ dan $|B^{1/2}\Sigma^{-1}B^{1/2}| = \prod_{i=1}^{p} \eta_i$

Kita dapat menuliskan : $|B^{1/2}\Sigma^{-1}B^{1/2}| = |B^{1/2}|\ |\Sigma^{-1}|\ |B^{1/2}| = |\Sigma^{-1}|\ |B^{1/2}|\ |B^{1/2}|$

$$= |\Sigma^{-1}|\ |B| = \frac{1}{|\Sigma|}|B|$$

atau

$$\frac{1}{|\Sigma|^b} = \frac{\left|B^{1/2}\Sigma^{-1}B^{1/2}\right|}{|B|} = \frac{\prod_{i=1}^{p}\eta_i}{|B|}$$

Kombinasikan hasil teras dan determinan,

$$\frac{1}{|\Sigma|^b} e^{-tr|\Sigma^{-1}B|/2} = \frac{\left(\prod_{i=1}^{p} \eta_i\right)^b}{|B|^b} e^{-\sum_{i=1}^{p} \eta_i/2} = \frac{1}{|B|^b} \prod_{i=1}^{p} \eta_i{}^b e^{-\eta_i/2}$$

Tetapi fungsi $\eta^b e^{-\eta/2}$ telah maksimum, dengan respek pada η, dari $(2b)^b e^{-b}$, terjadi pada η=2b. Pemilihan η_i = 2b, untuk setiap I, memberikan :

$$\frac{1}{|\Sigma|^b} e^{-tr(\Sigma^{-1})/2} \leq \frac{1}{|B|^b} (2b)^{pb} e^{-bp}$$

Batas atas yang unik dicapai ketika Σ = (1/2b)B. Pilihan ini mengakibatkan

$$B^{1/2} \Sigma^{-1} B^{1/2} = B^{1/2} (2b) B^{-1} B^{1/2} = (2b) \; I_{pxp}$$

dan

$$Tr [\Sigma^{-1}B] = tr [B^{1/2} \Sigma^{-1} B^{1/2}] = tr [(2b)I] = 2bp$$

Lebih jauh lagi,

$$\frac{1}{|\Sigma|} = \frac{\left|B^{1/2} \Sigma^{-1} B^{1/2}\right|}{|B|} = \frac{|(2b)I|}{|B|} = \frac{(2b)p}{|B|}$$

Substitusi untuk tr $[\Sigma^{-1}B]$ dan $1/|\Sigma|^b$ menghasilkan batas tesebut. Penduga maksimum likelihood untuk nilai μ dan Σ dinotasikan oleh $\hat{\mu}\, dan\, \hat{\Sigma}$ akan tergantung pada nilai $X_1, X_2, X_3, \ldots, X_n$ yang diobservasi melalui ringkasan statistik $\bar{X}$ dan S.

Misalkan $X_1, X_2, X_3, \ldots, X_n$ contoh acak dari populasi normal ganda dengan rata-rata μ dan covariance Σ. Maka $\hat{\mu} = \bar{X}$ dan

$\hat{\Sigma} = \frac{1}{n}\sum_{j=1}^{n}(X_j - \bar{X})(X_j - \bar{X})' = \frac{(n-1)}{n}S$ adalah penduga-penduga maksimum likelihood dari μ dan Σ. Nilai observasinya $\bar{X}$ dan $(1/n)\sum_{j=1}^{n}(X_j - \bar{X})(X_j - \bar{X})')$ disebut juga penduga maksimum likelihood dari μ dan Σ.

Bukti

Fungsi likelihood eksponen bagian dari fungsi multiplicative faktor –1/2 adalah

$$tr[\Sigma^{-1}(\sum_{j=1}^{n}(X_j - \bar{X})(X_j - \bar{X})')] + n(\bar{X} - \mu)'\Sigma^{-1}(\bar{X} - \mu) > 0 \text{ bila } \mu = \bar{X}.$$

Jadi likelihood dimaksimumkan dengan menduga $\hat{\mu} = \bar{X}$. Dianggap maksimumkan :

$$L(\hat{\mu}, \Sigma) = \frac{1}{(2\pi)^{np/2}|\Sigma|^{n/2}} e^{-tr[\Sigma^{-1}(\sum_{j=1}^{n}(X_j - \bar{X})(X_j - \bar{X})')]/2} \text{ over } \Sigma.$$

Dengan b = n/2 dan B = $\sum_{j=1}^{n}(X_j - \bar{X})(X_j - \bar{X})'$. Maksimum terjadi pada saat

$$\hat{\Sigma} = (1/n)\sum_{j=1}^{n}(X_j - \bar{X})(X_j - \bar{X})'.$$

3.11. Sebaran Penarikan Contoh dari $\overline{X}$ dan S

Dalam kasus univariate (p=1), kita tahu bahwa sebaran dari $\bar{x}$ adalah normal dengan rataan μ dan ragam σ^2/n. Hal ini juga berlaku untuk kasus multivariate (p≥2) dimana vektor rataan $\overline{X}$ menyebar normal dengan rataan μ dan matriks kovarian (1/n) Σ.

Ragam contoh $(n-1)S^2 = \sum_{j=1}^{n} (x_j - \bar{x})^2$ menyebar khi–kuadrat dengan derajat bebas n–1. Sebaran khi-kuadrat ini merupakan sebaran dari jumlah kuadrat peubah acak normal yang saling bebas, dimana $(n-1)S^2$ merupakan penjumlahan dari:

$$\sigma^2 (Z_1^2 + ... + Z_{n-1}^2) = (\sigma Z_1)^2 + ...+ (\sigma Z_{n-1})^2$$

lebih khusus lagi (σZ_i) adalah menyebar bebas dari N(0, σ^2). Ini merupakan bentuk umum sebaran percontohan dari unsur-unsur matriks kovarian.

Sebaran penarikan contoh dari matriks kovarian disebut sebaran Wishart,

W_m (. | Σ) = sebaran Wishart dengan db = m

= sebaran dari $\Sigma Z_j Z_j'$

dimana Z_j menyebar saling bebas yang merupakan bagian dari Np (0, Σ) maka kesimpulan sebaran penarikan contoh adalah :

Ambil x1, x2, ..., xn sebagai contoh acak yang berukuran n dari sebuah populasi normal ganda p dengan rataan μ dan matriks kovarian Σ, kemudian :
1. $\overline{X}$ adalah menyebar Np (μ, (1/n) Σ)
2. (n-1)S adalah menyebar Wishart dengan db = n – 1
3. $\overline{X}$ dan S adalah bebas.

Karena Σ tidak diketahui, sebaran $\bar{X}$ tidak dapat digunakan langsung untuk membuat turunan μ. Sebagaimana S memberikan informasi bebas tentang Σ dan distribusi S tidak bergantung pada μ.

Hal-hal yang diperlukan dari sebaran Wishart :

1. Jika A_1 menyebar W_{m1} $(A_1| \Sigma)$ bebas dari A_2, yang menyebar W_{m2} $(A_2| \Sigma)$, maka $A_1 + A_2$ menyebar W_{m1+m2} $(A_1 + A_2| \Sigma$
2. Jika A menyebar W_m $(A| \Sigma)$, maka CAC' menyebar W_m (CAC'|CΣC').

Fungsi kepekatan tidak ada jika ukuran contoh n tidak lebih besar dari jumlah peubah (p). Jika ada maka sebaran Wishartnya adalah :

$$W_{n-1}\langle A|\Sigma\rangle = \frac{|A|^{(n-p-2)/2} e^{-trA\Sigma-1/2}}{\Sigma^{p(n-1)/2} \prod^{p(p-1)/4} |\Sigma|^{(n-1)/2} \prod_{i=1}^{p} \Gamma[1/2(n-1)]}$$

dimana matriks A definit positif dan Γ (.) adalah fungsi Gamma.

3.12. Latihan

1. Angaplah vector peubah acak X menyebar normal ganda dengan $N_\theta(\underline{\mu}, \Sigma)$, dengan μ'=[1, -5, 1] dan

$$\Sigma = \begin{bmatrix} 1 & 5 & 1 \\ 5 & 3 & 2 \\ 1 & 2 & 2 \end{bmatrix}$$

   a. Carilah sebaran dari X_1-2X_2+X_3
   b. Carilah vector a, sehingga $X_1 - X_2 \; dan \, X_2 - a'\begin{bmatrix} X_1 \\ X_3 \end{bmatrix}$ saling bebas

2. Misalkan Y menyebar normal ganda dengan $N_\theta(\underline{\mu}, \Sigma)$, dengan μ'=[2, -3, 4] dan

$$\Sigma = \begin{bmatrix} 4 & -3 & 0 \\ -3 & 6 & 0 \\ 0 & 0 & 5 \end{bmatrix}$$

Manakah dari peubah-peubah berikut yang saling bebas?

a. y_1 dan y_2

b. y_1 dan y_3

c. y_2 dan y_3

d. (y_1, y_2) dan y_3

e. (y_1, y_3) dan y_2

4

4. Inferensia Vektor Nilai Tengah (*Inference of mean vector*)

4.1. Pendahuluan

Analisis statistika dapat dikategorikan menjadi dua yaitu statistika deskriptif dan statistika inferensia sebagian besar analisis statistika berkaitan dengan inferensia, yaitu menarik kesimpulan yang valid berdasarkan informasi dari sampel atau contah.

Sebelumnya telah kita pelajari mengenai inferensia nilai tengah dari suatu populasi. Pada makalah ini akan dibahas tentang inferensia tentang vektor nilai tengah dari sebaran normal ganda. Dalam menyelesaikannya kita menganalogikan dengan sebaran normal tunggal yang telah kita pelajari.

4.2. Test hipotesis vektor rataan

Suatu nilai μ_0 dianggap sebagai dugaan dari rataan populasi. Dari sisi pengujian dapat dituliskan sebagai berikut :

$$H_0 : \boldsymbol{\mu} = \boldsymbol{\mu_0} \text{ dan } H_1 : \boldsymbol{\mu} \neq \boldsymbol{\mu_0}$$

Dimana H_0 adalah hipotesis awal dan H_1 adalah hipotesis alternatif. Jika X1, X2,…, Xn menotasikan contoh acak dari sebuah populasi normal, statistik uji yang dipakai adalah :

Statistik uji di atas memiliki distribusi t-student dengan derajat bebas n-1. Kita menolak H0, yaitu menyatakan μ_0 bukan nilai yang tepat untuk menduga μ jika nilai |t| amatan lebih besar dari sebuah titik yang sudah ditentukan dari sebaran t dengan derajat bebas n-1.

Menolak H0 ketika |t| besar setara dengan menolak H0 jika kuadratnyacukup besar, yaitu

Peubah T^2 adalah jarak kuadrat dari rataan contoh ke nilai uji μ_0. Unit-unit jarak diekspresikan dengan simpangan baku dugaan dari rataan contoh.

Apabila rataan contoh dan ragam sudah didapat dari perhitungan, pengujiannya adalah menolak H0 dan menerima H1 pada taraf nyata α jika :

Dimana tn-1((/2) menotasikan persentil 100((/2) persentil atas dari sebaran t dengan derajat bebas n-1.

Jika H0 tidak ditolak, maka (0 adalah nilai yang dapat diterima untuk menduga rataan populasi normal.

Untuk pengujian hipotesis diatas, selang kepercayaan untuk μ adalah:

Selang kepercayaan memuat semua nilai μ_0 yang akan diterima pada pengujian hipotesis.

Dalam menentukan apakah suatu vektor μ_0 p x q adalah nilai yang dapat diterima untuk rataan dari sebuah sebaran normal peubah ganda.

Generalisasi dari jarak dengan analogi peubah ganda adalah

Jika nilai T^2 terlalu besar, dimana nilai rataan populasi terlalu jauh dari nilai rataan dugaan, maka hipotesis awal (H_0) ditolak.

T^2 sebarannya adalah (n-1)p/(n-p) $F_{p,n-p}$ dimana $F_{p,n-p}$ menotasikan sebuah peubah acak dengan sebaran F yang memiliki p dan n-p peubah bebas.

Pada uji hipotesis $H_0 : \mu = \mu_0$ lawan $H_1 : \mu \neq \mu_0$ pada taraf nyata α, tolak H_0 jika :

Sebuah daerah kepercayaan 100(1-α)% dari sebuah rataan sebaran normal dimensi p ditentukan oleh semua μ adalah :

Karena hal di atas dapat dianalogikan dengan uji hipotesis, dapat dilihat bahwa daerah kepercayaan di atas mengandung semua vektor μ_0 dimana uji T^2 tidak akan menolak H_0 pada taraf nyata α.

Untuk dimensi yang lebih dari 4, daerah kepercayaan gabungan untuk μ tidak dapat digambarkan, tetapi panjang relatif dari sumbu-sumbu dari selang kepercayaan dapat dihitung. Hal tersebut ditentukan dari akar dan vektor ciri dari S. Arah dan panjang dari sumbu daerah kepercayaan ditentukan dari :

$$\sqrt{\lambda_i}\, c/\sqrt{n} = \sqrt{\lambda_i}\sqrt{p(n-1)F_{p,n-p}(\alpha)/n(n-p)}$$

Unit sepanjang vektor ciri $\mathbf{e_i}$. Berawal dari pusat rataan, sumbu-sumbu dari elipsoid kepercayaan adalah :

$$\pm\sqrt{\lambda_i}\sqrt{\frac{p(n-1)}{n(n-p)}F_{p,n-p}(\alpha)}\, e_i$$

Selang kepercayaan untuk peubah ke-p adalah :

$$\bar{x}_p - \sqrt{\frac{p(n-1)}{(n-p)}F_{p,n-p}(\alpha)}\sqrt{\frac{s_{pp}}{n}} \le \mu_p \le \bar{x}_p + \sqrt{\frac{p(n-1)}{(n-p)}F_{p,n-p}(\alpha)}\sqrt{\frac{s_{pp}}{n}}$$

Metode alternatif untuk perbandingan berganda adalah dengan menggunakan metode Bonferroni

Dimana m=banyaknya selang kepercayaan selang kepercayaan yang akan dibuat.

$$\overline{x_i} \pm t_{n-1}\left(\frac{\alpha}{2m}\right)\sqrt{\frac{s_{ii}}{n}}$$

Teladan:

Data diperoleh dari 87 mahasiswa, X_1= sosial scien and history , X_2 = verbal dan X_3 = science, dari data diketahui

$$\bar{X} = \begin{bmatrix} 526.586 \\ 54.690 \\ 25.92 \end{bmatrix}, \quad \mu_0 = \begin{bmatrix} 500 \\ 50 \\ 30 \end{bmatrix}, \quad S = \begin{bmatrix} 5808.06 & 597.84 & 222.92 \\ 597.84 & 126.05 & 22.96 \\ 22.92 & 22.96 & 22.94 \end{bmatrix},$$

$$S^{-1} = \begin{bmatrix} 0.0004387 & -0.0015949 & -0.0026660 \\ -0.0015949 & 0.0154997 & -0.0000143 \\ -0.0026660 & -0.0000143 & 0.0694983 \end{bmatrix}$$

$T^2{}_{hitung}$ = n $(x - \mu_0)' S^{-1} (x-\mu_0)$ = 228.266

Hipotesa yang ingin diuji adalah:

$H_0 : \mu = (500\ 50\ 30)'$

$H_1 : \mu \neq (500\ 50\ 30)'$

Nilai $T^2{}_{hitung}$, yang sudah diperoleh selanjutnya dibandingkan dengan nilai T^2 tabel, yang dapat diperoleh dari konversi dengan nilai F, yaitu:

$$T^2 = \frac{(n-1)p}{n-p} F_{p,(n-p)} = \frac{(87-1)3}{87-3} F_{3,(87-3)} = \frac{258}{84} F_{3,84} = 3{,}07(2{,}71) = 8{,}33$$

T^2 tabel < T^2 hitung, dengan demikian dapat disimpulkan tolak H_0.

4.3. Aplikasi SAS

Berikut adalah contoh program untuk pengujian dua vektor rataan menggunakan PROC IML.

```
data X1X2;
input x1 x2;
cards;
3.7    48.5
5.7    65.1
3.8    47.2
3.2    53.2
3.1    55.5
4.6    36.1
2.4    24.8
7.2    33.1
6.7    47.4
5.4    54.1
3.9    36.9
4.5    58.8
3.5    27.8
4.5    40.2
1.5    13.5
8.5    56.4
4.5    71.6
6.5    52.8
4.1    44.1
5.5    40.9
;

proc iml;
TITLE1 'uji dua vektor';
USE X1X2;
read all var {x1 x2 } into xx;
u0={4,50};
xr=xx[:,];
*hitung matriks peragam;
n=nrow(xx);
satu=repeat(1,n,1);
I0=i(n)-(1/n)#(satu*satu`);
s=(1/(n-1))#(xx`*i0*xx);
is=inv(s);
*hitung t2;
T2=n#(xr`-u0)`*is*(xr`-u0);
print T2;
```

```
* nilai kritis;
p=ncol(xx);
alpha=0.1;
F = FINV(1-alpha,p,n-p);
nk=((n-1)#p/(n-p))#F;
print n p f nk;
if T2>nk then
   print 'T2 > nilai kritis';
else
   print 'T2 < nilai kritis';

*buat titik kordinat ellips;
call eigen(val,vec,s);
print val vec;
*sumbu mayor;
mayor=sqrt(val[1])*sqrt(nk/n);
minor=sqrt(val[2])*sqrt(nk/n);
print mayor minor;

quit;
```

4.4. Latihan

1. Suatu perusahaan memproduksi sabun dengan dua metode yang berbeda untuk memilih metode produksi yang terbaik dilakukan pengamatan terhadap 50 buah sabun untuk setiap metode. Karaketeristik sabun yang diamati antara lain X_1 = lather dan X2=mildness . Ringkasan datanya diperoleh sebagai berikut :

$$\bar{x}_1 = \begin{bmatrix} 8.3 \\ 4.0 \end{bmatrix}, S_1 = \begin{bmatrix} 3 & 1 \\ 1 & 6 \end{bmatrix}, \bar{x}_2 = \begin{bmatrix} 10.0 \\ 3.90 \end{bmatrix}, S_2 = \begin{bmatrix} 2 & 1 \\ 1 & 4 \end{bmatrix}$$

   Asumsi : Semua kondisi di luar metode dikondisikan homogen

   a. Buatlah selang kepercayaan 95% simultan untuk beda nilai tengah kedua metode. Lengkapi dengan gambar elipsnya
   b. Ujilah apakah vector rata-rata penggunaan energi kedua wilayah sama pada taraf nyata 5%

2. Empat puluh sembilan (49) orang dari Indonesia dan 47 orang dari Malaysia yang diukur karakteristik personalnya. Beberapa variable yang mewakili karakteristik personal adalah :

   X1 = daya tangkap terhadap informasi lisan

   X2 = daya tangkap terhadap informasi arah

   X3 = daya tangkap terhadap informasi aritmatika

   X4 = daya tangkap terhadap informasi gambar

   Data rata-rata yang ditunjukkan oleh dua grup adalah sebagai berikut :

Variable	**Grup**	
	Indonesia	**Malaysia**
Rata-rata X1	12.57	8.75
Rata-rata X2	9.57	5.33
Rata-rata X3	11.49	8.50
Rata-rata X4	7.97	4.75

   Matriks Kovariannya adalah sebagai berikut

$$S = \begin{bmatrix} 11.26 & 9.41 & 7.15 & 3.38 \\ & 13.53 & 7.38 & 2.51 \\ & & 11.58 & 2.62 \\ & & & 5.81 \end{bmatrix}$$

Kebalikan matriks kovariance adalah sebagai berikut :

$$S^{-1} = \begin{bmatrix} 0.2598 & -0.1365 & -0.0587 & -0.0659 \\ & 0.1867 & -0.0383 & 0.0162 \\ & & 0.1510 & -0.0173 \\ & & & 0.2111 \end{bmatrix}$$

Apakah karakteristik personal orang indonesia sama dengan karakteristik personal orang malaysia. Gunakan alpha 5 %.

3. Suatu percobaan terhadap 11 rokok yang diukur kadar tar dan kadar nikotin. Sebelum diberi filter rokok-rokok tersebut diisap dan diukur kadar tar dan kadar nikotinnya pada asap yang tersedot. Selanjutnya pada rokok-rokok tersebut dipasang filter dan rokok diisap dengan kondisi asap melewati filter tersebut. Asap rokok yang melewati filter diukur kadar tar dan nikotinnya. Data adalah sebagai berikut :

No	Rokok sebelum dipasang filter		Rokok setelah dipasang filter	
	Kadar Tar	Kadar Nikotin	Kadar Tar	Kadar Nikotin
1	28	13	6	23
2	25	15	6	27
3	35	29	8	44
4	36	22	18	64
5	44	64	34	75
6	15	31	11	30
7	42	30	28	26
8	54	64	71	124
9	29	20	33	30
10	34	56	43	54
11	39	21	20	14

Apakah kadar tar dan nikotin rokok sama antara tanpa filter dengan diberi filter ? Gunakan alpha 5 %.

5

5.Manova (*Multivariate Analysis of Variance*)

5.1. Pendahuluan

Pada dasarnya analisis ragam peubah ganda (*multivariate analysis of variance* atau *manova*) merupakan pengembangan lebih lanjut dari analisis ragam satu peubah (*anova*). Berikut ini beberapa perbedaan dari kedua analisis tersebut:

Analisis Ragam Satu Peubah (Anova)	Analisis Ragam Peubah Ganda (Manova)
Hanya mengkaji pengaruh berbagai per-cobaan yang dilakukan terhadap respon tunggal (satu buah peubah respon).	Mengkaji pengaruh dari berbagai perla-kuan yang dicobakan terhadap respons ganda (lebih dari satu peubah respon).
Ketergantungan di antara peubah respon tidak menjadi perhatian utama karena pada dasarnya dianggap bahwa peubah-peubah respons itu saling bebas satu sama lain , sehingga pengkajian struktur keragaman hanya dilakukan terhadap setiap peubah respons secara terpisah.	Mempertinbangkan adanya ketergan-tungan antar peubah-peubah respons, sehingga baik dugunakan untuk pengkajian pengaruh dari berbagai perlakuan terhadap lebih dari satu respons.

5.2. Analisis Ragam Peubah Ganda Satu Arah (*One-way Manova*)

Analisis ragam peubah ganda satu arah merupakan pengembangan dari analisis ragam satu peubah satu arah. Oleh sebab itu, sebelum membahas analisis ragam peubah ganda satu arah perlu dikemukakan tentang analisis ragam satu peubah satu arah.

Model yang sering digunakan dalam analisis ragam satu arah adalah model *Rancangan Acak Lengkap* (*RAL*).

$$Y_{ij} = \mu + \tau_i + \varepsilon_{ij}$$

di mana: i = 1, 2, ... , t. j = 1, 2, ... , n_{i}.

Y_{ij} = nilai pengamatan (respon tunggal) dari ulangan ke-j yang memperoleh perlakuan ke-i.

μ = nilai rataan umum.

τ_i = pengaruh dari perlakuan ke-i terhadap respon.

ε_{ij} = pengaruh galat yang timbul pada ulangan ke-j dan perlakuan ke-i.

Asumsi yang diperlukan pada analisis ragam satu peubah adalah :

$\sum \tau_i = 0$ dan $\varepsilon_{ij} \sim N(0, \sigma^2)$

Dalam analisis ragam satu peubah, pengujian hipotesis dilakukan terhadap adanya pengaruh perlakuan , di mana:

$H_0 : \tau_1 = \tau_2 = \cdots = \tau_t = 0$

atau tidak terdapat pengaruh dari semua perlakuan yang diterapkan terhadap respon yang diamati.

$H_1 : \exists\, i, \tau_i \neq 0,\ i = 1,2,\cdots,t$

atau paling sedikit ada satu perlakuan yang mempengaruhi respon yang diamati.

Jika kita ingin mengkaji pengaruh dari t buah perlakuan terhadap p buah respon secara serempak, di mana p > 1, maka penelitian itu dapat dianalisis dengan analisis ragam peubah ganda satu arah.

Model umum dari analisis ragam peubah ganda satu arah adalah

$$Y_{ijk} = \mu_k + \tau_{ik} + \varepsilon_{ijk}$$

di mana: i =1, 2, … , t, j=1, 2, … , n_i, k=1, 2, … , p.

Y_{ijk} = nilai pengamatan dari respon ke-k dan ulangan ke-j yang memperoleh perlakuan ke-i.

μ_k = nilai rataan umum dari respon ke-k.

τ_{ik} = pengaruh dari perlakuan ke-i terhadap respon ke-k.

ε_{ijk} = pengaruh galat yang timbul pada respon ke-k dari ulangan ke-j dan perlakuan ke-i.

Asumsi yang dibutuhkan dalam analisis ragam peubah ganda satu arah adalah nilai-nilai galat bersifat bebas dan menyebar normal ganda dengan vektor nilai rata-rata **0** dan matriks peragam Σ (ε_{ijk} ~ N_p (**0**, Σ)).

Ilustrasi:

Seorang peneliti bidang kedokteran melakukan percobaan untuk meneliti hubungan di antara aktifitas metabolik di antara kelinci-kelinci

percobaan dan daya tahan terhadap kuman tuberculosis (tbc). Peneliti menetapkan 4 perlakuan sebagai berikut:

P_1 = kontrol (tidak divaksinasi)

P_2 = diinfeksi (ditularkan) kuman tbc selama aktifitas metabolik rendah.

P_3 = diinfeksi (ditularkan) kuman tbc selama aktifitas metabolik tinggi.

P_4 = diinfeksi (ditularkan) kuman tbc selama aktifitas metabolik normal, tetapi terlebih dahulu diirradiasi dengan 400 rontgens.

Perlakuan P_1 dan P_2 diulang sebanyak 7 kali ($n_1 = n_2 = 7$), perlakuan P_3 diulang 5 kali ($n_3 = 5$), dan P_4 diulang sebanyak 2 kali ($n_4 = 2$). Peubah respon yang diamati ada 2 yaitu:

Y_1 = banyaknya basil yang hidup per tubercle formed (mm).

Y_2 = banyaknya basil yang hidup per tubercle size (mm).

Data hasil pengamatan seperti pada tabel di bawah ini.

Banyaknya Basil yang Hidup per *Tubercle Formed* (Y_1) dan *Tubercle Size* (Y_2) dalam mm

Ulangan	P_1		P_2		P_3		P_4	
	Y_1	Y_2	Y_1	Y_2	Y_1	Y_2	Y_1	Y_2
1	24.0	3.5	7.4	3.5	16.4	3.2	25.1	2.7
2	13.3	3.5	13.2	3.0	24.0	2.5	5.9	2.3
3	12.2	4.0	8.5	3.0	53.0	1.5		
4	14.0	4.0	10.1	3.0	32.7	2.6		
5	22.2	3.6	9.3	2.0	42.8	2.0		
6	16.1	4.3	8.5	2.5				
7	27.9	5.2	4.3	1.5				
Total	129.7	28.1	61.3	18.5	168.9	11.8	31.0	5.0
Rata-rata	18.5286	4.0143	8.7571	2.6428	33.7800	2.3600	15.5000	2.5000

Analisis ragam peubah ganda satu arah untuk data di atas dilakukan sebagai berikut:

- *Perhitungan Faktor Koreksi (FK) untuk untuk respon Y_1 dan Y_2*

FK untuk respon Y_1

$$FK_{11} = \frac{(Y_{1\bullet\bullet})^2}{n} = \frac{(390.9)^2}{21} = 7276.3243$$

FK untuk respon Y_2

$$FK_{22} = \frac{(Y_{2\bullet\bullet})^2}{n} = \frac{(63.4)^2}{21} = 191.4076$$

FK untuk respon Y_1 dan Y_2

$$FK_{12} = \frac{(Y_{1\bullet\bullet})(Y_{2\bullet\bullet})}{n} = \frac{(390.9)(63.4)}{21} = 1180.1457$$

- *Perhitungan Jumlah Kuadrat Total Terkoreksi (JKT) dan Jumlah Hasil Kali Total Terkoreksi (JHKT) untuk respon Y_1 dan Y_2*

JKT untuk respon Y_1

$$JKT_{11} = \sum_{i=1}^{4}\sum_{j=1}^{n_i} Y_{1ij}^2 - FK_{11} = 3152.2657$$

JKT untuk respon Y_2

$$JKT_{22} = \sum_{i=1}^{4}\sum_{j=1}^{n_i} Y_{2ij}^2 - FK_{22} = 17.4124$$

JHKT untuk respon Y_1 dan Y_2

$$JHKT_{12} = \sum_{i=1}^{4}\sum_{j=1}^{n_i} Y_{1ij} \cdot Y_{2ij} - FK_{12} = -39.0257$$

Dari hasil perhitungan untuk komponen total, maka dapat dibentuk suatu matriks **T** yang memuat elemen-elemen JKT dan JHKT yang berkaitan dengan respon Y_1 dan Y_2, sebagai berikut:

$$\mathbf{T} = \begin{pmatrix} 3152.2657 & -39.0257 \\ -39.0257 & 17.4124 \end{pmatrix}$$

- *Perhitungan Jumlah Kuadrat Perlakuan Terkoreksi (JKP) dan Jumlah Hasil Kali Perlakuan Terkoreksi (JHKP) untuk respon Y_1 dan Y_2*

JKP untuk respon Y_1

$$JKP_{11} = \sum_{i=1}^{4} \frac{Y_{1i\bullet}^2}{n_i} - FK_{11} = 1849.5862$$

JKP untuk respon Y2

$$JKP_{22} = \sum_{i=1}^{4} \frac{Y_{2i\bullet}^2}{n_i} - FK_{22} = 10.6346$$

JHKP untuk respon Y_1 dan Y_2

$$JHKP_{12} = \sum_{i=1}^{4} \frac{Y_{1i\bullet} \cdot Y_{2i\bullet}}{n_i} - FK_{12} = -21.3810$$

Dari hasil perhitungan untuk komponen perlakuan, maka dapat dibentuk suatu matriks **P** yang memuat elemen-elemen JKP dan JHKP yang berkaitan dengan respon Y_1 dan Y_2, sebagai berikut:

$$\mathbf{P} = \begin{pmatrix} 1849.5862 & -21.3810 \\ -21.3810 & 10.6346 \end{pmatrix}$$

- *Perhitungan Jumlah Kuadrat Galat (JKG) dan Jumlah Hasil Kali Galat (JHKG) untuk respon Y_1 dan Y_2*

JKG untuk respon Y_1

$$JKG_{11} = JKT_{11} - JKP_{11} = 1302.6795$$

JKG untuk respon Y_2

$$JKG_{22} = JKT_{22} - JKP_{22} = 6.7778$$

JHKG untuk respon Y_1 dan Y_2

$$JHKG_{12} = JHKT_{12} - JHKP_{12} = -17.6447$$

Dari hasil perhitungan untuk komponen galat, maka dapat dibentuk suatu matriks **G** yang memuat elemen-elemen JKG dan JHKG yang berkaitan dengan respon Y_1 dan Y_2, sebagai berikut:

$$\mathbf{G} = \begin{pmatrix} 1302.6795 & -17.6447 \\ -17.6447 & 6.7778 \end{pmatrix}$$

Hasil perhitungan yang diperoleh sebelumnya dapat dirangkum dalam suatu tabel analisis ragam peubah ganda satu arah (*One-way Manova*) seperti berikut:

Tabel Analisis Ragam Peubah Ganda Satu Arah(*One-way Manova*)

Sumber Keragaman	Derajat Bebas (db)	JK dan JHK
Perlakuan (P)	3	$\mathbf{P} = \begin{pmatrix} 1849.5862 & -21.3810 \\ -21.3810 & 10.6346 \end{pmatrix}$
Galat (G)	17	$\mathbf{G} = \begin{pmatrix} 1302.6795 & -17.6447 \\ -17.6447 & 6.7778 \end{pmatrix}$
Total (T)	20	$\mathbf{T} = \begin{pmatrix} 3152.2657 & -39.0257 \\ -39.0257 & 17.4124 \end{pmatrix}$

Untuk menguji hipotesis yang telah dikemukakan sebelumnya, maka kita dapat menggunakan uji *Lambda-Wilks* (Λ-*Wilks*), sebagai berikut:

$$\Lambda = \frac{|\mathbf{G}|}{|\mathbf{G}+\mathbf{P}|} = \frac{|\mathbf{G}|}{|\mathbf{T}|}$$

di mana:

$|\mathbf{G}|$ = determinan dari matriks galat (**G**)

$|\mathbf{T}|$ = determinan dari matriks total (**T**)

Selanjutnya besaran Λ yang dihitung dari rumus di atas dibandingkan dengan tabel distribusi U dengan kaidah keputusan sebagai berikut:

$$\text{jika } \Lambda \begin{cases} > U^{\alpha}_{p;db_P;db_G} \text{ maka terima } H_0 \\ \leq U^{\alpha}_{p;db_P;db_G} \text{ maka tolak } H_0 \end{cases}$$

di mana:

p = banyaknya peubah respon yang diamati.

db_P = derajat bebas perlakuan.

db_G = derajat bebas galat.

Untuk kasus di atas maka berdasarkan hasil perhitungan yang telah dirangkum dalam tabel *Manova* didapat $|\mathbf{G}| = 8517.9657$ dan $|\mathbf{T}| = 53365.5060$, sehingga besaran *Lambda-Wilks* dapat dihitung sebagai berikut:

$$\Lambda = \frac{|\mathbf{G}|}{|\mathbf{T}|} = \frac{8517.9657}{53365.5060} = 0.1596$$

Dalam kasus ini, p=2, db_P=3, dan db_G=17. Jika kita menetapkan taraf nyata pengujian hipotesis adalah α = 0.01, maka dari tabel distribusi U didapat bahwa $U^{0.01}_{2;3;17}$ = 0.370654. Karena Λ = 0.1596 < $U^{0.01}_{2;3;17}$ = 0.370654 maka sesuai dengan kaidah keputusan di atas maka H_0 ditolak. Dengan demikian dapat disimpulkan bahwa terdapat perbedaan daya hidup basil di antara keempat perlakuan, atau vektor nilai rata-rata berbeda di antara empat perlakuan yang dicobakan.

Kita juga dapat melakukan transformasi dari besaran Λ ke besaran statistik F sehingga dapat dibandingkan dengan tabel distribusi F. Bentuk transformasi dari nilai Λ ke nilai F untuk berbagai kombinasi jumlah peubah (p) dan derajat bebas perlakuan (db_P) disajikan dalam tabel.

Dalam kasus di atas, p=2, db_P=3, hal ini berarti sesuai dengan kriteria transformasi F untuk p=2 dan $db_P \geq 1$, sehingga transformasi dari Λ ke F dapat dilakukan sebagai berikut:

$$F = \left(\frac{1-\sqrt{\Lambda}}{\sqrt{\Lambda}}\right)\left(\frac{db_G - 1}{db_P}\right) = 8.017$$

Selanjutnya besaran F ini dibandingkan dengan nilai dari tabel F dengan derajat bebas 2 db_P;2(db_G –1). Jika kita menetapkan α=0.01, maka $F_{0.01;6;32}$ = 3.434. Karena besaran F = 8.017 > $F_{0.01;6;32}$ = 3.434, maka kita

menolak H_0 pada taraf α=0.01. Dengan demikian terlihat bahwa hasil pengujian berdasarkan statistik F sama dengan statistik *Lambda-Wilks*.

Tabel Transformasi dari Λ ke F

Parameter		Transformasi F	Derajat Bebas
P	db_P		
1	≥ 1	$\left(\frac{1-\Lambda}{\Lambda}\right)\left(\frac{db_G}{db_P}\right)$	$db_P;db_G$
2	≥ 1	$\left(\frac{1-\sqrt{\Lambda}}{\sqrt{\Lambda}}\right)\left(\frac{db_G-1}{db_P}\right)$	2 db_P;2(db_G –1)
≥ 1	1	$\left(\frac{1-\Lambda}{\Lambda}\right)\left(\frac{db_P+db_G-p}{p}\right)$	p;(db_P + db_G - p)
≥ 1	2	$\left(\frac{1-\sqrt{\Lambda}}{\sqrt{\Lambda}}\right)\left(\frac{db_P+db_G-p-1}{p}\right)$	2p;2(db_P + db_G – p –1)

5.3. Analisis Ragam Peubah Ganda Dua Arah (Two-way Manova)

Seperti halnya analisis ragam peubah ganda satu arah, analisis ragam peubah ganda dua arah juga merupakan pengembangan lebih lanjut dari analisis ragam satu peubah dua arah (*Two-way Anova*).

Model yang sering digunakan dalam analisis ragam satu peubah dua arah adalah *Rancangan Acak Kelompok* (*RAK*).

$$Y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij}$$

di mana: *i* = 1, 2, … , *t*; *j* = 1, 2, … , *r*

Y_{ij} = nilai pengamatan (respon tunggal) dari kelompok ke-j yang memperoleh perlakuan ke-i.

μ = nilai rataan umum.

τ_i = pengaruh dari perlakuan ke-i terhadap respon.

β_j = pengaruh dari kelompok ke-j terhadap respon.

ε_{ij} = pengaruh galat yang timbul pada kelompok ke-j yang memperoleh perlakuan ke-i.

Dalam analisis ragam peubah ganda dua arah, model yang digunakan adalah :

$$Y_{ijk} = \mu_k + \tau_{ik} + \beta_{jk} + \varepsilon_{ijk}$$

di mana: *i* = 1, 2, … , *t*; *j* = 1, 2, … , *r*; *k* = 1, 2, … , *p*.

Y_{ijk} = nilai pengamatan dari respon ke-k dan kelompok ke-j yang memperoleh perlakuan ke-i.

μ_k = nilai rataan umum dari respon ke-k.

τ_{ik} = pengaruh dari perlakuan ke-i terhadap respon ke-k.

β_{jk} = pengaruh dari kelompok ke-j terhadap respon ke-k.

ε_{ijk} = pengaruh galat yang timbul pada respon ke-k dari kelompok ke-j yang memperoleh perlakuan ke-i.

Asumsi yang diperlukan adalah nilai-nilai galat bersifat bebas dan menyebar normal ganda dengan vektor nilai rata-rata **0** dan matriks peragam Σ (ε_{ijk} ~ N_p (**0**, Σ)).

Dalam analisis ragam peubah ganda dua arah, hipotesis yang diuji adalah :

$$H_0 : \begin{pmatrix} \tau_{11} \\ \tau_{12} \\ \cdots \\ \tau_{1p} \end{pmatrix} = \begin{pmatrix} \tau_{21} \\ \tau_{22} \\ \cdots \\ \tau_{2p} \end{pmatrix} = \cdots = \begin{pmatrix} \tau_{t1} \\ \tau_{t2} \\ \cdots \\ \tau_{tp} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \cdots \\ 0 \end{pmatrix}$$

atau tidak ada pengaruh perlakuan terhadap respons yang diamati.

$$H_1 : \exists\, i, \tau_{ik} \neq 0, i = 1,2,\cdots,t \text{ dan } k = 1,2,\cdots,p$$

atau paling sedikit ada satu perlakuan yang mempengaruhi respon pengamatan.

Ilustrasi:

Dari data yang mempunyai 3 peubah respon Y_1, Y_2, dan Y_3, yang diamati dari 5 perlakuan yaitu P_1, P_2, P_3, P_4, dan P_5, yang dikenakan pada 4 kelompok yaitu K_1, K_2, K_3, dan K_4, seperti di bawah ini:

Data untuk Analisis Ragam Peubah Ganda Dua Arah

Perlakuan	Kelompok 1			Kelompok 2			Kelompok 3			Kelompok 4			Total		
	Y_1	Y_2	Y_3	Y_1	Y_2	Y_3	Y_1	Y_2	Y_3	Y_1	Y_2	Y_3	Y_1	Y_2	Y_3
P_1	96	10	725	142	16	700	122	13	655	111	13	680	417	52	2760
P_2	102	15	695	106	10	710	95	14	705	93	12	680	396	51	2790
P_3	109	15	690	113	15	690	101	14	680	100	19	685	423	63	2745
P_4	103	17	680	97	16	690	99	13	730	135	12	670	434	58	2770
P_5	98	17	680	97	14	695	105	16	680	86	22	710	386	69	2765
Total	508	74	3470	555	71	3485	522	70	3450	525	78	3425	2110	293	13830

Dapat disusun tabel *Manova* sesuai dengan cara yang telah dijelaskan sebelumnya seperti berikut ini:

Tabel Analisis Ragam Peubah Ganda Dua Arah

Sumber Keragaman	Derajat Bebas	JK dan JHK
Kelompok (K)	3	$\mathbf{K} = \begin{pmatrix} 234.60 & -14.10 & 127.00 \\ -14.10 & 7.75 & -36.50 \\ 127.00 & -36.50 & 405.00 \end{pmatrix}$
Perlakuan (P)	4	$\mathbf{P} = \begin{pmatrix} 1129.50 & -125.75 & -213.75 \\ -125.75 & 57.30 & -62.00 \\ -213.75 & -62.00 & 267.50 \end{pmatrix}$
Galat (G)	12	$\mathbf{G} = \begin{pmatrix} 2258.90 & -24.65 & -1658.25 \\ -24.65 & 91.50 & 4.00 \\ -1658.25 & 4.00 & 5532.50 \end{pmatrix}$
Total (T)	19	$\mathbf{T} = \begin{pmatrix} 3623.00 & -164.50 & -1745.00 \\ -164.50 & 156.55 & -94.50 \\ -1745.00 & -94.50 & 6205.00 \end{pmatrix}$

Pengujian hipotesis dalam analisis ragam peubah ganda dua arah dapat menggunakan statistik *Lambda-Wilks* sebagai berikut:

$$\Lambda = \frac{|\mathbf{G}|}{|\mathbf{G}+\mathbf{P}|} = \frac{8.888319 \times 10^8}{2.227617 \times 10^9} = 0.399006$$

Selanjutnya besaran Λ dibandingkan dengan nilai U dari tabel distribusi U pada taraf nyata α dengan p=3, db_P=4, db_G=12. Jika kita menetapkan α=0.05, maka dari tabel distribusi U diperoleh $U^{0.05}_{3;4;12}$=0.168939. Karena Λ=0.399006 > $U^{0.05}_{3;4;12}$=0.168939, maka kita menerima H_0, dan menyatakan bahwa berdasarkan data yang ada kita belum dapat menolak hipotesis kesamaan pengaruh perlakuan.

5.4. Aplikasi SAS

ANOVA adalah bagian dari prosedur statistika untuk membahas perbedaan rataan dari beberapa populasi menggunakan data contoh yang ditarik dari populasi yang bersesuaian. Ada banyak jenis Anova, mulai dari model satu arah (one-way), model dua arah dengan dan tanpa interaksi, rancangan dengan pengaruh acak, rancangan tak lengkap dan sebagainya. Anova hanya melibatkan satu peubah respon. Manova adalah perluasan konsep dan teknik ANOVA pada situasi ada beberapa peubah respon. Berikut disajikan ilustrasi penggunaan SAS untuk MANOVA.

5.4.1. One-way Manova

Seorang peneliti bidang kedokteran melakukan percobaan untuk meneliti hubungan di antara aktifitas metabolik di antara kelinci-kelinci

percobaan dan daya tahan terhadap kuman tuberculosis (tbc). Peneliti menetapkan 4 perlakuan sebagai berikut:

P1 = kontrol (tidak divaksinasi)

P2 = diinfeksi (ditularkan) kuman tbc selama aktifitas metabolik rendah.

P3 = diinfeksi (ditularkan) kuman tbc selama aktifitas metabolik tinggi.

P4 = diinfeksi (ditularkan) kuman tbc selama aktifitas metabolik normal, tetapi terlebih dahulu diradiasi dengan 400 rontgens.

Perlakuan P1 dan P2 diulang sebanyak 7 kali (n1 = n2 = 7), perlakuan P3 diulang 5 kali (n3 = 5), dan P4 diulang sebanyak 2 kali (n4 = 2). Peubah respon yang diamati ada 2 yaitu:

Y1 = banyaknya basil yang hidup per tubercle formed (mm).

Y2 = banyaknya basil yang hidup per tube rcle size (mm).

Datanya adalah sebagai berikut :

Ulangan	P1		P2		P3		P4	
	Y1	Y2	Y1	Y2	Y1	Y2	Y1	Y2
1	24.0	3.5	7.4	3.5	16.4	3.2	25.1	2.7
2	13.3	3.5	13.2	3.0	24.0	2.5	5.9	2.3
3	12.2	4.0	8.5	3.0	53.0	1.5		
4	14.0	4.0	10.1	3.0	32.7	2.6		
5	22.2	3.6	9.3	2.0	42.8	2.0		
6	16.1	4.3	8.5	2.5				
7	27.9	5.2	4.3	1.5				

Data tersebut terlebih dahulu dimasukkan dalam SAS data set sebagai berikut :

```
data basilus;
      title1 "Data Basilus";
      input Perlakuan $ Y1 Y2 ;
      datalines;
P1      24       3.5
P1      13.3     3.5
P1      12.2     4
P1      14       4
P1      22.2     3.6
P1      16.1     4.3
P1      27.9     5.2
P2      7.4      3.5
P2      13.2     3
P2      8.5      3
P2      10.1     3
P2      9.3      2
P2      8.5      2.5
P2      4.3      1.5
P3      16.4     3.2
P3      24       2.5
P3      53       1.5
P3      32.7     2.6
P3      42.8     2
P4      25.1     2.7
P4      5.9      2.3
;

 proc glm data=basilus;
      class Perlakuan;
      model Y1 Y2 = Perlakuan;
      manova h=_all_ / printe printh;
    run;
```

Dengan menggunakan PROC GLM diperoleh hasil sebagai berikut :

```
                              Data Basilus

                           The GLM Procedure
                        Class Level Information

                      Class          Levels    Values
                    Perlakuan             4     P1 P2 P3 P4

                  Number of Observations Read          21
                  Number of Observations Used          21

Dependent Variable: Y1
                                      Sum of
Source                    DF         Squares      Mean Square       F Value      Pr > F
Model                      3     1849.586286       616.528762          8.05      0.0015
Error                     17    1302.679429        76.628202
Corrected Total           20     3152.265714


                 R-Square      Coeff Var       Root MSE       Y1 Mean
                 0.586748       47.02707       8.753754      18.61429

        Source            DF       Type I SS     Mean Square    F Value    Pr > F
         Perlakuan          3     1849.586286     616.528762       8.05    0.0015

          Source         DF     Type III SS     Mean Square    F Value     Pr > F
```

```
        Perlakuan     3   1849.586286     616.528762      8.05   0.0015

Dependent Variable: Y2
                                      Sum of
Source                  DF     Squares       Mean Square       F Value     Pr > F
Model                    3  10.63466667      3.54488889         8.89     0.0009
Error                   17    6.77771429      0.39868908
Corrected Total         20   17.41238095

               R-Square     Coeff Var      Root MSE       Y2 Mean
               0.610753      20.91449      0.631418      3.019048

        Source         DF     Type I SS   Mean Square  F Value   Pr > F
        Perlakuan      3   10.63466667    3.54488889     8.89    0.0009

        Source         DF  Type III SS    Mean Square   F Value  Pr > F
        Perlakuan      3   10.63466667    3.54488889     8.89    0.0009

                        E = Error SSCP Matrix
                         Y1                Y2
                  Y1      1302.6794286          -17.644
                  Y2           -17.644      6.7777142857

    Partial Correlation Coefficients from the Error SSCP Matrix / Prob > |r|

                  DF = 17            Y1            Y2

                  Y1           1.000000      -0.187775
                                 0.4556
                  Y2          -0.187775       1.000000
                                 0.4556

                  H = Type III SSCP Matrix for Perlakuan

                         Y1                Y2

                  Y1      1849.5862857      -21.38171429
                  Y2      -21.38171429      10.634666667

          Characteristic Roots and Vectors of: E Inverse * H, where
                  H = Type III SSCP Matrix for Perlakuan
                        E = Error SSCP Matrix

         Characteristic               Characteristic Vector  V'EV=1
               Root        Percent         Y1              Y2

           1.59856728      53.12      0.01118710      0.38173632
           1.41098630      46.88      0.02589504     -0.08492416

    MANOVA Test Criteria and F Approximations for the Hypothesis of No Overall
                          Perlakuan Effect
                  H = Type III SSCP Matrix for Perlakuan
                        E = Error SSCP Matrix

                          S=2    M=0    N=7

Statistic                         Value  F Value  Num DF  Den DF  Pr > F

Wilks' Lambda            0.15961411   8.02     6       32    <.0001
Pillai's Trace           1.20040454   8.51     6       34    <.0001
Hotelling-Lawley Trace   3.00955358   7.82     6    19.652   0.0002
Roy's Greatest Root      1.59856728   9.06     3       17    0.0008
```

```
NOTE: F Statistic for Roy's Greatest Root is an upper bound.
NOTE: F Statistic for Wilks' Lambda is exact.
```

Berdasarkan keluaran SAS di atas diperoleh nilai statistic Wilks' Lambda dengan p-value < .05. Ini menunjukkan bahwa cukup bukti untuk menolak Ho yang berarti bahwa minimal terdapat sepasang perlakuan yang memberikan pengaruh yang berbeda terhadap respon.

5.4.2. Two Way Manova

Berikut merupakan data hasil penelitian yang melibatkan lima perlakuan dengan empat kelompok.

perlakuan	kelompok1			Kelompok 2			Kelompok 3			Kelompok 4		
	Y1	Y2	Y3	Y1	Y2	Y3	Y1	Y2	Y3	Y1	Y2	Y3
p1	96	10	725	142	16	700	122	13	655	111	13	680
p2	102	15	695	106	10	710	95	14	705	93	12	680
p3	109	15	690	113	15	690	101	14	680	100	19	685
p4	103	17	680	97	16	690	99	13	730	135	12	670
p5	98	17	680	97	14	695	105	16	680	86	22	710
total	508	74	3470	555	71	3485	522	70	3450	525	78	3425

Untuk mengetahui apakah ada pengaruh perlakuan terhadap respon, dengan menggunakan SAS perintahnya sebagai berikut:

```
data rekaan;
      title1 "Data Rekaan";
      input Kelompok $ Perlakuan $ Y1 Y2 Y3;
      datalines;
K1     P1      96      10      725
K1     P2      102     15      695
K1     P3      109     15      690
K1     P4      103     17      680
K1     P5      98      17      680
K2     P1      142     16      700
K2     P2      106     10      710
K2     P3      113     15      690
K2     P4      97      16      690
K2     P5      97      14      695
```

```
K3     P1      122     13      655
K3     P2      95      14      705
K3     P3      101     14      680
K3     P4      99      13      730
K3     P5      105     16      680
K4     P1      111     13      680
K4     P2      93      12      680
K4     P3      100     19      685
K4     P4      135     12      670
K4     P5      86      22      710
   ;
   proc glm data=rekaan;
      class Kelompok Perlakuan;
      model Y1 Y2 Y3 = Kelompok Perlakuan;
      manova h=_all_ / printe printh;
   run;
```

Keluaran yang dihasilkan dari PROC GLM di atas adalah sebagai berikut:

```
                              Data Rekaan

                           The GLM Procedure

                        Class Level Information

                    Class           Levels    Values

                  Kelompok               4    K1 K2 K3 K4

                  Perlakuan              5    P1 P2 P3 P4 P5


                 Number of Observations Read          20
                 Number of Observations Used          20


Dependent Variable: Y1

                                        Sum of
Source              DF         Squares     Mean Square    F Value    Pr > F

Model                7     1364.100000      194.871429       1.04    0.4562

Error               12     2258.900000      188.241667

Corrected Total     19     3623.000000


                R-Square     Coeff Var      Root MSE       Y1 Mean

                0.376511      13.00485      13.72012      105.5000


Source              DF       Type I SS     Mean Square    F Value    Pr > F

Kelompok            3      234.600000      78.200000       0.42    0.7451
Perlakuan           4     1129.500000     282.375000       1.50    0.2634
```

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Kelompok	3	234.600000	78.200000	0.42	0.7451
Perlakuan	4	1129.500000	282.375000	1.50	0.2634

Dependent Variable: Y2

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	7	65.0500000	9.2928571	1.22	0.3638
Error	12	91.5000000	7.6250000		
Corrected Total	19	156.5500000			

R-Square	Coeff Var	Root MSE	Y2 Mean
0.415522	18.84874	2.761340	14.65000

Source	DF	Type I SS	Mean Square	F Value	Pr > F
Kelompok	3	7.75000000	2.58333333	0.34	0.7977
Perlakuan	4	57.30000000	14.32500000	1.88	0.1790

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Kelompok	3	7.75000000	2.58333333	0.34	0.7977
Perlakuan	4	57.30000000	14.32500000	1.88	0.1790

Dependent Variable: Y3

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	7	672.500000	96.071429	0.21	0.9768
Error	12	5532.500000	461.041667		
Corrected Total	19	6205.000000			

R-Square	Coeff Var	Root MSE	Y3 Mean
0.108380	3.105117	21.47188	691.5000

Source	DF	Type I SS	Mean Square	F Value	Pr > F
Kelompok	3	405.0000000	135.0000000	0.29	0.8298
Perlakuan	4	267.5000000	66.8750000	0.15	0.9617

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Kelompok	3	405.0000000	135.0000000	0.29	0.8298
Perlakuan	4	267.5000000	66.8750000	0.15	0.9617

The GLM Procedure
Multivariate Analysis of Variance

E = Error SSCP Matrix

	Y1	Y2	Y3
Y1	2258.9	-24.65	-1658.25

```
Y2           -24.65              91.5                  4
Y3         -1658.25                 4             5532.5


   Partial Correlation Coefficients from the Error SSCP Matrix / Prob > |r|

             DF = 12            Y1            Y2            Y3

             Y1           1.000000     -0.054220     -0.469073
                                          0.8604        0.1059

             Y2          -0.054220      1.000000      0.005622
                            0.8604                      0.9855

             Y3          -0.469073      0.005622      1.000000
                            0.1059        0.9855


                              The GLM Procedure
                       Multivariate Analysis of Variance

                     H = Type III SSCP Matrix for Kelompok

                    Y1                 Y2                 Y3

Y1            234.6              -14.1                127
Y2            -14.1               7.75              -36.5
Y3              127              -36.5                405


          Characteristic Roots and Vectors of: E Inverse * H, where
                     H = Type III SSCP Matrix for Kelompok
                             E = Error SSCP Matrix

         Characteristic               Characteristic Vector  V'EV=1
       Root    Percent              Y1              Y2              Y3

0.26074591      74.69     0.01925198     -0.03643290      0.01169745
0.06529416      18.70     0.01317611      0.08381796     -0.00042845
0.02306344       6.61    -0.00502711      0.05112645      0.00973873


  MANOVA Test Criteria and F Approximations for the Hypothesis of No Overall
                               Kelompok Effect
                     H = Type III SSCP Matrix for Kelompok
                             E = Error SSCP Matrix

                             S=3    M=-0.5    N=4

Statistic                 Value   F Value   Num DF   Den DF   Pr > F

Wilks' Lambda          0.72778034    0.38      9    24.488    0.9338
Pillai's Trace         0.29065442    0.43      9        36    0.9105
Hotelling-Lawley Trace 0.34910352    0.37      9    12.866    0.9310
Roy's Greatest Root    0.26074591    1.04      3        12    0.4089

        NOTE: F Statistic for Roy's Greatest Root is an upper bound.


                    H = Type III SSCP Matrix for Perlakuan
```

```
                    Y1                  Y2                  Y3

Y1            1129.5             -125.75             -213.75
Y2           -125.75                57.3                 -62
Y3           -213.75                 -62               267.5
```

```
             Characteristic Roots and Vectors of: E Inverse * H, where
                      H = Type III SSCP Matrix for Perlakuan
                               E = Error SSCP Matrix


          Characteristic               Characteristic Vector  V'EV=1
        Root        Percent         Y1               Y2               Y3

0.91280518     74.84      0.01564533     -0.07482267      0.00527950
0.29514853     24.20      0.01707089      0.06744138      0.00032600
0.01165071      0.96      0.00577310      0.02863476      0.01427856

  MANOVA Test Criteria and F Approximations for the Hypothesis of No Overall
                              Perlakuan Effect
                      H = Type III SSCP Matrix for Perlakuan
                               E = Error SSCP Matrix

                                S=3    M=0    N=4

Statistic                       Value  F Value  Num DF  Den DF  Pr > F

Wilks' Lambda              0.39900569    0.93     12    26.749    0.5366
Pillai's Trace             0.71661193    0.94     12      36      0.5186
Hotelling-Lawley Trace     1.21960442    0.95     12    13.692    0.5293
Roy's Greatest Root        0.91280518    2.74      4        12    0.0789

         NOTE: F Statistic for Roy's Greatest Root is an upper bound.
```

5.5. Latihan

1. Suatu percobaan dilakukan menggunakan 8 jenis pupuk. Pupuk-pupuk tersebut kemudian disebar pada petak-petak lahan yang ditanami padi. Karena lahan tidak homogen maka lahan di blok menjadi 8 blok. Setiap blok ada 8 petak. Pengacakan 8 perlakuan dilakukan pada setiap blok. Pada saat panen diukur bobot biji dan bobot serasak per petak. Matriks jumlah kuadrat adalah sebagai berikut :

P (Matriks Perlakuan)= $\begin{bmatrix} 12496.8 & -6786.8 \\ -6786.8 & 32985.0 \end{bmatrix}$

B (Matriks Blok) = $\begin{bmatrix} 86045.8 & 56073.6 \\ 56073.6 & 75841.5 \end{bmatrix}$

E (Matriks Galat) = $\begin{bmatrix} 136792.6 & 58549.0 \\ 58549.0 & 71496.1 \end{bmatrix}$

Apakah ada pengaruh pemupukan terhadap respon ? (gunakan alpha 5 %).

2. Suatu percobaan dilakukan untuk mengetahui perbedaan tiga varietas jagung. Data respon yang diambil antara lain Y_1 = Produksi per hektar, dan Y_2 = bobot/1000 butir. Rancangan lingkuan yang digunakan adalah rancangan acak lengkap. Datanya diperoleh sebagai berikut :

Perlakuan	**Ulangan**	**Y_1**	**Y_2**
Varietas – 1	1	6	7
	2	5	9
Varietas - 2	1	4	6
	2	6	6
	3	4	7
Varietas - 3	1	5	4
	2	6	4

a. Tuliskan model linearnya, lengkap dengan keterangannya!
b. Hitunglah vector rataan untuk setiap perlakuan.
c. Hitunglah matriks jumlah kuadrat dan hasil kali silang dari perlakuan (B), galat (W), dan Total (T)
d. Lakukan pengujian pada taraf α = 5% untuk mengetahui apakah ketiga varietas memiliki respon yang berbeda. Gunakan uji Wilks' Lambda!
e. Apa kesimpulan Anda?

3. Sebuah observasi dengan 3 respon untuk 3 perlakuan. Hasil dari vektor observasi $\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$

Perlakuan 1 : $\begin{bmatrix} 6 \\ 7 \\ 9 \end{bmatrix}$ $\begin{bmatrix} 5 \\ 9 \\ 10 \end{bmatrix}$ $\begin{bmatrix} 8 \\ 6 \\ 9 \end{bmatrix}$

Perlakuan 2 : $\begin{bmatrix} 3 \\ 3 \\ 5 \end{bmatrix}$ $\begin{bmatrix} 1 \\ 6 \\ 4 \end{bmatrix}$ $\begin{bmatrix} 2 \\ 3 \\ 5 \end{bmatrix}$

Perlakuan 3: $\begin{bmatrix} 2 \\ 3 \\ 5 \end{bmatrix}$ $\begin{bmatrix} 5 \\ 1 \\ 4 \end{bmatrix}$ $\begin{bmatrix} 2 \\ 3 \\ 6 \end{bmatrix}$

a. Buatlah tabel MANOVA
b. Hitunglah nilai Wilks Lambda untuk menguji pengaruh perlakuan dengan a=0.01.

4. Perhatikan observasi dengan respon x_1 dan x_2. Terdapat dua faktor dengan tanpa perlakuan pada tabel berikut.

		Faktor 2			
		Level 1	Level 2	Level 3	Level 4
faktor 1	Level 1	$\begin{bmatrix} 6 \\ 8 \end{bmatrix}$	$\begin{bmatrix} 4 \\ 6 \end{bmatrix}$	$\begin{bmatrix} 8 \\ 12 \end{bmatrix}$	$\begin{bmatrix} 2 \\ 6 \end{bmatrix}$
	Level 2	$\begin{bmatrix} 3 \\ 8 \end{bmatrix}$	$\begin{bmatrix} -3 \\ 2 \end{bmatrix}$	$\begin{bmatrix} 4 \\ 3 \end{bmatrix}$	$\begin{bmatrix} -4 \\ 3 \end{bmatrix}$
	Level 3	$\begin{bmatrix} -3 \\ 2 \end{bmatrix}$	$\begin{bmatrix} -4 \\ -5 \end{bmatrix}$	$\begin{bmatrix} 3 \\ -3 \end{bmatrix}$	$\begin{bmatrix} -4 \\ -6 \end{bmatrix}$

		Faktor 2			
		Level 1	Level 2	Level 3	Level 4
faktor 1	Level 1	$\begin{bmatrix} 14 \\ 8 \end{bmatrix}$	$\begin{bmatrix} 6 \\ 2 \end{bmatrix}$	$\begin{bmatrix} 8 \\ 2 \end{bmatrix}$	$\begin{bmatrix} 16 \\ -4 \end{bmatrix}$
	Level 2	$\begin{bmatrix} 1 \\ 6 \end{bmatrix}$	$\begin{bmatrix} 5 \\ 12 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 5 \end{bmatrix}$	$\begin{bmatrix} 2 \\ 7 \end{bmatrix}$

Level 3	$\begin{bmatrix} 3 \\ -2 \end{bmatrix}$	$\begin{bmatrix} -2 \\ 7 \end{bmatrix}$	$\begin{bmatrix} -11 \\ 1 \end{bmatrix}$	$\begin{bmatrix} -6 \\ 6 \end{bmatrix}$

a. Tuliskanlah model liniernya lengkap dengan keterangan!
b. Hitunglah vektor rataan untuk setiap perlakuan!
c. Lakukanlah uji Wilks Lambda untuk mengetahui apakah masing-masing taraf pada masing-masing faktor memberikan respon yang berbeda!

5. Seorang kepala sekolah melakukan penelitian untuk mendapatkan metode pembelajaran yang terbaik. Terdapat empat metode yang diterapkan di beberapa kelompok kelas. Hasil yang didapat untuk mata pelajaran Matematika (MAT), Bahasa Inggris (BI) dan IPA adalah sebagai berikut:

METODE	KELOMPOK 1			KELOMPOK 2			KELOMPOK 3			KELOMPOK 4		
	MAT	BI	IPA	MAT	BI	IPA	MAT	BI	IPA	MAT	BI	IPA
Metode 1	70	60	70	85	65	75	90	70	80	95	75	85
Metode 2	75	65	75	89	69	80	94	76	85	96	80	90
Metode 3	89	85	81	91	73	82	99	81	90	100	85	97
Metode 4	92	82	85	100	80	90	105	84	93	110	90	101

a. tuliskan model liniear lengkap dengan keterangan
b. ujilah apakah metode yang dilakukan memberikan espon yang berbeda?
c. Apakah pengelompokan yang dilakukan berdasarkan kelas tersebut efektif?

6

6. Analisis Profil (*Profile Analysis*)

6.1. Pendahuluan

Analisis profil berkenaan dengan situasi dimana serangkaian *p* perlakuaan yang dikenakan terhadap dua populasi (kelompok) atau lebih. Dan biasanya didalam melakukan percobaan dengan situasi seperti itu, sering kali kita ingin mengetahui pengaruh perlakuan yang satu dengan yang lainnya untuk setiap populasi (kelompok). Salah satu metode yang dapat digunakan untuk menyelesaikan masalah tersebut adalah dengan menggunakan analisis profile (*profile analysis*).

Menurut *Morisson* (1991) analisis profile merupakan suatu bagian dari pengujian hipotesis terhadap nilai tengah dari peubah ganda (*multivariate*) dengan menggunakan prinsip grafik. Dengan demikian untuk mengetahui perkiraan tentang kemiripan profile baik profile antar perlakuan maupun antar kelompok yang dinyatakan dengan kesejajaran itu, dapat kita lihat dari grafik plot antara nilai rataan tiap-tiap perlakuan untuk setiap kelompok (populasi). Tetapi hanya dengan melihat grafik saja tidaklah cukup, kita juga perlu untuk mengetahui seberapa besar arti kesejajaran (kemiripan) dari populasi itu. Untuk itulah diperlukan serangkaian uji-uji yang berkaitan dengan hipotesis itu.

Untuk melakukan analisis profile ini diperlukan asumsi-asumsi sebagai berikut :

1. Setiap perlakuan untuk kelompok (populasi) yang berbeda bersifat saling bebas satu dengan lainnya.
2. Seluruh respon dari peubah-peubahnya harus dinyatakan dengan satuan yang sama agar dapat dibandingkan dan dijumlahkan .
3. Nilai galatnya menyebar multinormal dengan rataan 0 dan ragam σ.

Ada tiga pertanyaan atau hipotesis yang akan di uji didalam anlisis profile, yaitu :

1. Apakah profile –profile itu sejajar ?
   atau setara dengan H_0:μ_{1l}-μ_{1l-1}=μ_{2l}-μ_{2l-1} untuk l= 2,3,4,…p .
2. Jika profile itu sejajar, apakah profile-profile itu saling berhimpit?.
   atau setara dengan H_0 : μ_{1l}=μ_{2l} untuk l= 1,2,3 …p.
3. Jika profile-profile itu saling berhimpit apakah profile-profile itu memiliki besaran yang sama ?
   Atau setara dengan H_0 : μ_{11}=μ_{22}=…=μ_{1p}=μ_{21}=μ_{22}=…=μ_{2p}. (sejajar dengan sumbu X (datar)).

Hipotesis pertama berkaitan dengan interaksi (pengaruh) antar kelompok perlakuan. Jika sejajar, maka interaksi (pengaruh) antar perlakuan tersebut tidak ada. Sedangkan hipotesis kedua berkaitan dengan hipotesis kesamaan pengaruh setiap perlakuan pada tiap kelompok. Jika berhimpit, maka nilai tengah untuk masing-masing perlakuan tiap kelompok akan sama. Dan hipotesis ketiga berkaitan dengan kesemua perlakuan itu mempunyai nilai tengah yang sama untuk setiap kelompok (populasi).

Ketiga hipotesis diatas tersebut haruslah diuji secara berurutan. Artinya bahwa jika hipotesis pertama (mengenai kesejajaran), setelah diuji ternyata di tolak, maka uji-uji untuk hipotesis dua (keberhimpitan) dan tiga (kehorisontalan) tidak akan berlaku lagi.

6.2. Pengujian Hipotesis

Apabila analisis profile dinotasikan dalam persamaan matriks, maka model umumnya adalah sebagai berikut :

$$\begin{bmatrix} y_{11} \\ . \\ y_{1n_1} \\ y_{21} \\ . \\ y_{2n_2} \\ . \\ y_{i1} \\ . \\ y_{in_i} \end{bmatrix} = \begin{bmatrix} 1...0......0 \\ .... \\ 1...0......0 \\ 0...1......0 \\ .... \\ 0...1......0 \\ .... \\ 0...0......1 \\ .... \\ 0...0......1 \end{bmatrix} \begin{bmatrix} \mu_{11} & \mu_{12} & \cdots & \mu_{1p} \\ \mu_{21} & \mu_{22} & \cdots & \mu_{2p} \\ . & . & \cdots & . \\ \mu_{i1} & \mu_{i2} & \cdots & \mu_{ip} \end{bmatrix} + \begin{bmatrix} e_{11} \\ .. \\ e_{1n_1} \\ e_{21} \\ .. \\ e_{2n_2} \\ .. \\ e_{i1} \\ .. \\ e_{in_i} \end{bmatrix}$$

Atau dapat juga ditulis

$$\mathbf{Y}=\mathbf{XB}+\mathbf{E},$$

dengan **X** adalah matriks rancangan berdimensi (N x I), **B** matriks parameter berdimensi (I x p), dan **E** matriks galat berdimensi (N x p). Sedangkan **Y** merupakan matriks peubah tak bebas berdimensi (N x p). Dengan p = jumlah peubah tak bebas, I= jumlah perlakuan (populasi), n_l= jumlah pengamatan pada perlakuan ke-l dan N= jumlah total pengamatan.

6.2.1. Uji Kesejajaran (*Parallel Test*)

Bentuk umum hipotesisnya :

$$H_{01} = \begin{bmatrix} \mu_{11}-\mu \\ \mu_{12}-\mu_{13} \\ \vdots \\ \mu_{1(p-1)}-\mu_{1p} \end{bmatrix} = \begin{bmatrix} \mu_{21}-\mu_{22} \\ \mu_{22}-\mu_{23} \\ \vdots \\ \mu_{2(p-1)}-\mu_{2p} \end{bmatrix} = \ldots = \begin{bmatrix} \mu_{i1}-\mu_{i2} \\ \mu_{i2}-\mu_{i3} \\ \vdots \\ \mu_{i(p-1)}-\mu_{ip} \end{bmatrix}$$

Uji kesejajaran untuk *dua populasi* yang menyebar normal dapat dituliskan sebagai berikut:

H_{01} : $C\mu_1$ = $C\mu_2$ dimana C merupakan matriks kontras sedemikian sehingga membuat persamaan seperti pada bentuk umum hipotesis kesejajaran diatas.

$$C_{(p-1)xp} = \begin{bmatrix} -1 & 1 & 0 & ..... \\ 0 & -1 & 1 & ..... \\ .. & .. & .. & ..... \\ 0 & 0 & 0 & ..... \end{bmatrix}$$

Untuk contoh bebas dari dua populasi (perlakuan), maka kita dapat membuat nilai rataan untuk tiap-tiap peubahnya sehingga akan kita dapatkan rataan dari populasi 1 x_1 dan rataan dari populasi 2 x_2 .

Dan pengujian nya adalah sebagai berikut

$$T^2 = (x_1 - x_2)'C'\left[(\frac{1}{n_1} + \frac{1}{n_2})CSC'\right]^{-1} C(x_1 - x_2)$$

$$dengan$$

$$c^2 = \frac{(n_1 + n_2 - 2)(p-1)}{n_1 + n_2 - p} F_{p-1, n_1+n_2-p}(\alpha)$$

dengan S adalah matriks koragam (Covarian) dari peubah-peubahnya.

Hipotesis nol ditolak jika nilai dari T^2 > c^2 . Dengan nilai dari c^2 nya tergantung dari nilai tabel sebaran F dengan db1= p-1 dan db2=n1 + n2 – p pada (α)

6.2.2. Uji Keberhimpitan (*Coincident Test*)

Bentuk umum dari hipotesisnya adalah :

$$H_{02} = \begin{bmatrix} \mu_{11} \\ \mu_{12} \\ : \\ \mu_{1p} \end{bmatrix} = \begin{bmatrix} \mu_{21} \\ \mu_{22} \\ : \\ \mu_{2p} \end{bmatrix} = \ldots = \begin{bmatrix} \mu_{I1} \\ \mu_{i2} \\ : \\ \mu_{ip} \end{bmatrix}$$

Atau dengan kata lain, profile akan saling berhimpit apabila total dari nilai tataan tiap-tiap populasi

$$\mu_{11}+\mu_{12}+\ldots+\mu_{1p}=\mu_{21}+\mu_{22}+\ldots+\mu_{2p}=\ldots\ldots=\mu_{i1}+\ldots..+\mu_{ip} \ .$$

dan untuk *dua populasi* yang normal maka bentuk hipotesis nolnya adalah

$$\mathbf{H_0 : \quad 1'\mu_1 = 1'\mu_2} \ .$$

Pengujian hipotesis ini baru dapat dilakukan setelah uji pada kesejajaran dapat di terima. Statistik uji untuk pengujian hipotesis keberhimpitan dapat ditulis sebagai :

$$T^2 = 1'(\overline{x_1} - \overline{x_2})\left[(\frac{1}{n_1} + \frac{1}{n_2})1'S1\right]1'(\overline{x_1} - \overline{x_2})$$

$$T^2 = \left(\frac{1'(\overline{x_1} - \overline{x_2})}{\sqrt{(\frac{1}{n_1} + \frac{1}{n_2})1'S1}}\right)^2$$

Untuk kaidah pengambilan keputusannya adalah kita akan menolak Hipotesis nol apabila nilai dari statistik uji T^2 tersebut diatas > $t^2_{n1+n2-2}$ $(\alpha/2)$ (distribusi t tabel dengan db=n1+n2-p pada level (α) dikuadratkan).

Atau kita juga akan menolak hipotesis nol apabila $T^2 > F_{p-1,n1+n2-p}(\alpha)$ (distribusi F tabel dengan db1=p-1 dan db2=n1+n2-p pada level (α))

6.2.3. Uji Kesamaan (*Level Test*)

Apabila profil-profil tersebut berhimpit (hipotesis nol keberhimpitan diterima), maka seluruh observasi tersebut berasal dari populasi normal yang sama. Maka langkah selanjutnya adalah apakah seluruh peubahnya tersebut memiliki nilai rataan yang sama .

Ketika kesejajaran dan keberhimpitan dapat diterima, maka vektor rataan μ (dari dua populasi normal) dapat diduga dengan menggunakan n1+n2 observasi (pengamatan) sebagai berikut:

$$\bar{x} = \frac{\sum_{j=1}^{n1} x_{1j} + \sum_{j=1}^{n2} x_{2j}}{n_1 + n_2} = \frac{n_1}{n_1 + n_2}\bar{x}_1 + \frac{n_2}{n_1 + n_2}\bar{x}_2$$

Jika profile itu sama (se-level), maka $\mu_1 = \mu_2 = \ldots = \mu_p$.

Bentuk hipotesis nolnya dapat kita tuliskan sebagai :

$$H_{03} = \begin{bmatrix} \mu_{11} \\ \mu_{12} \\ \vdots \\ \mu_{1p} \end{bmatrix} = \begin{bmatrix} \mu_{21} \\ \mu_{22} \\ \vdots \\ \mu_{2p} \end{bmatrix} = \ldots = \begin{bmatrix} \mu_{i1} \\ \mu_{i2} \\ \vdots \\ \mu_{ip} \end{bmatrix}$$

Atau dapat juga dituliskan sebagai : **H_{03} : $C\mu=0$** .

Statistik uji yang digunakan adalah :

$$F = (n_1 + n_2)\bar{x}'C'\left[CSC'\right]^{-1}C\bar{x}$$

Hipotesis nol ditolak jika statistik uji diatas F > dari $F_{p-1,n1+n2-p}$ (α) (lebih besar dari nilai distribusi F tabel dengan derajat kebebasan db1 = p-1 dan db2= n1+n2-p pada taraf (level) pengujian (α)).

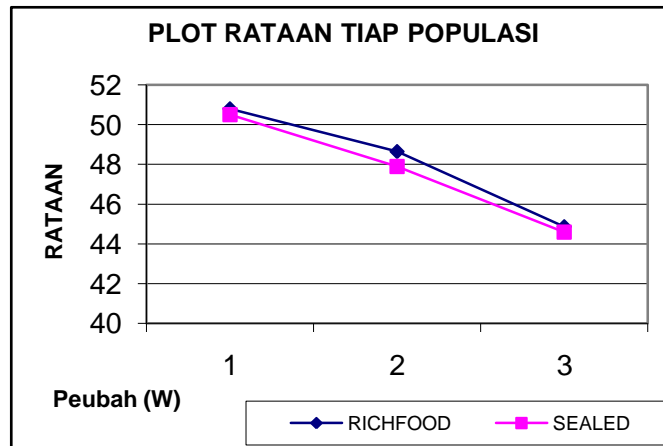
6.2.4. Ilustrasi

Jurusan Gizi dam makanan di Virginia Polytecnic Institute and State University pada tahun 1976 melakukan penelitian mengenai *"Vitamin C Retention in Reconstituted Frozen Orange Juice"* untuk mengetahui bahwa kadar asam askorbat menentukan kestabilan vitamin C.

Untuk tujuan tersebut, dilakukan penelitian pada dua jenis sari air jeruk yaitu merk Richfood dan Sealeda yang masing-masing diulang sebanyak empat kali. Dua sari air jeruk tersebut disimpan dalam lemari es selama satu minggu. Selama selang waktu tersebut diukur kadar asam askorbatnya dalam satuan miligram(mg) per liter pada 3 waktu yang berbeda yaitu waktu ke-0, waktu ke-3, dan waktu ke-7. Data yang diperoleh dari percobaan tersebut adalah:

Merk	W0	W3	W7
Richfood	52.6	49.4	42.7
Sealed	56.0	48.8	49.2
Richfood	54.2	49.2	48.8
Sealed	48.0	50.3	44.0
Richfood	49.8	42.8	40.4
Sealed	49.6	44.0	42.0
Richfood	46.6	53.2	47.6
Sealed	48.4	48.5	43.2

6.2.4.1. Pembahasan dengan Metode Grafik

Menurut Morisson (1991) , pada dasarnya analisis profile merupakan suatu bagian dari uji hipotesis terhadap nilai tengah dari peubah ganda dengan menggunakan prinsip grafik. Dari data di atas, maka kita memiliki 2 kelompok yaitu Richfood dan Sealed, serta tiga peubah yaitu W-0, W-3, dan W-7. Untuk memulai analisis dengan menggunakan metode grafik ini,

kita hanya membutuhkan nilai rataan tiap-tiap peubah didalam setiap populasi. Kemudian setelah mendapatkan nilainya, dibuatlah grafik plot untuk setiap populasi. Didapat rataan tiap-tiap peubah untuk Richfood = [58,8 48,65 44,87] dan untuk populasi Sealed = [50,5 47,9 44,6]. Kemudian dari nilai nilai ini didapat plot sebagai berikut:

Dari grafik di atas sudah dapat digambarkan bahwa profil untuk setiap kelompok tersebut sejajar dan berhimpit. Seberapa jauh nilai dari kesejajaran dan keberhimpitannya bersifat relatif, karena dengan menggunakan metode grafik pengujiannya masih bersifat kasar. Namun demikian, hasil yang diperoleh dari pendekatan secara grafik seharusnya tidak jauh berbeda dengan hasil pengujian melalui uji-uji analisis profile.

6.2.4.2. Pembahas dengan Metode Uji-uji Analisis Profil

Uji akan dilakukan pada taraf nyata 0.05. Diketahui bahwa rataan untuk tiap populasi Xr=[50.8, 48.65, 44.87] dan Xs=[50.5, 47.9, 44.6], sehingga didapat matriks selisih rataan kedua populasi ter sebut yaitu X=[0.3, 0.75, 0.27]. Sedangkan nilai dari matriks koragam peubah-peubahnya nya dan nilai matrik C yang digunakan berukuran 2x3 yaitu:

	$w0$	$w3$	$w7$
$w0$	10.7629	3.1471	4.9579
$w3$	3.1471	15.4964	9.3632
$w7$	4.9579	9.3632	11.1398

$$C = \begin{bmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix}$$

$$A = \left(\frac{1}{6} \begin{bmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} 10.76 & 3.14 & 4.96 \\ 3.14 & 15.5 & 9.36 \\ 4.96 & 9.36 & 11.14 \end{bmatrix} \begin{bmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix}^t \right)^{-1}$$

sehingga

$$T^2 = \begin{bmatrix} 0.3 \\ 0.75 \\ 0.27 \end{bmatrix}^t \begin{bmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix}^t A \begin{bmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} 0.3 \\ 0.75 \\ 0.27 \end{bmatrix}$$

Selanjutnya pengujian kesejajaran dapat dilakukan menggunakan software SAS IML dengan membuat makro SAS sesuai prossedur perhitungannya. Tahapan yang dilakukan dalam pembuatan makro SAS:

1. Hitung nilai T^2.

2. Hitung nilai c^2.

Dengan menggunakan data di atas, nilai T^2 = 0.1753. dan nilai c^2 dengan $F_{2,21}$ (0.05) = 3,47 dan n1=12 n2=12 serta p=3 yaitu :

$$c^2 = \frac{(12+12-2)(3-1)}{12+12-3} 3.47 = \frac{(22)(2)}{21} 3.47 = 7.27$$

Untuk mengambil keputusan, dilakukan perbandiingan antara nilai T^2 dan c^2. Karena nilai T^2 < c^2, maka dapat disimpulkan bahwa profile antara

kelompok (merk) tersebut adalah sejajar. Atau dengan kata lain bahwa seiring dengan bertambahnya waktu maka asam askorbat pada kedua kelompok tersebut mengalami penambahan dan pengurangan dengan proporsi yang sama (tidak ada perbedaan pengaruh perlakuan). Kesimpulan ini juga didukung oleh tampilan grafik diatas.

Selain menggunakan kedua pendekatan di atas, beberapa nilai statistik dapat juga digunakan sebagai dasar untuk pengambilan keputusan uji kesejajaran. Nilai statistik tersebut antara lain: wilks'lamda, pillai's trace, Hottelling-lawleytrace dan Roy's greatest root. Keempat nilai statistic tersebut dapat diperoleh menggunakan PROC GLM pada software SAS. Nilai ke empat statistik berdasarkan data di atas:

Statistic	Value	F	Num DF	Den DF	Pr > F
Wilks' Lambda	0.990898	0.023	2	5	0.9774
Pillai's Trace	0.009102	0.023	2	5	0.9774
Hotelling-Lawley Trace	0.009185	0.023	2	5	0.9774
Roy's Greatest Root	0.009185	0.023	2	5	0.9774

Dengan menggunakan α= 0.05, maka dapat disimpulkan profile antara kelompok (merk) tersebut adalah sejajar (nilai Pr>F lebih besar dari 0.05).

Setelah hipotesis mengenai kesejajaran diterima, maka langkah selanjutnya adalah uji mengenai keberhimpitan.

Didefinisikan matriks 1' =[1 1 1], statistik uji keberhimpitan adalah :

$$T^2 = \begin{bmatrix}1\\1\\1\end{bmatrix}^t \begin{bmatrix}0.3\\0.75\\0.27\end{bmatrix} \left(\frac{1}{6} \begin{bmatrix}1\\1\\1\end{bmatrix}^t \begin{bmatrix}10.76 & 3.14 & 4.96\\3.14 & 15.5 & 9.36\\4.96 & 9.36 & 11.14\end{bmatrix} \begin{bmatrix}1\\1\\1\end{bmatrix} \right)^{-1} \begin{bmatrix}1\\1\\1\end{bmatrix}^t \begin{bmatrix}0.3\\0.75\\0.27\end{bmatrix}$$

Dapat dihitung dengan menggunakan proc IML dalam SAS bahwa nilai T^2 =0.38.

Kaidah pengambilan keputusan dilakukan dengan membandingkan nilai T^2 dengan nilai F tabel dengan db1=1 dan db2=22 pada level 0.05 yang nilainya sama dengan 4.30. Karena nilai T^2 < F tabel, maka dapat disimpulkan bahwa profile kedua kelompok itu saling berhimpit. Hal ini menunjukkan bahwa kedua merk sari air jeruk tersebut memiliki kadar asam askorbat yang sama setelah disimpan selama seminggu.

Uji keberhimpitan dapat juga dilihat dari hasil analisis ragam untuk tiap-tiap waktu pengamatan :

Dependent Variable: W-0

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	0.18	0.18000000	0.01	0.9085
Error	6	75.16	12.52666667		
Corrected Total	7	75.34			

Dependent Variable: W-3

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	1.125	1.12500000	0.09	0.7784
Error	6	77.93	12.98833333		
Corrected Total	7	79.055			

Dependent Variable: W-7

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	0.15125	0.15125000	0.01	0.917
Error	6	77.8275	12.97125		
Corrected Total	7	77.97875			

Setelah mengetahui bahwa profile kedua merk sari air jeruk (kelompok) tersebut sejajar dan berhimpit, maka uji pada analisis profile yang ketiga yaitu mengenai kesamaan (rataan kandungan asam karbonat dari waktu

kewaktu adalah sama) dapat kita lakukan. Kita akan menemukan matriks rataan dari waktu kewaktu secara keseluruhan yaitu:

$$\overline{X} = \frac{12}{24}\begin{bmatrix} 50.8 \\ 48.65 \\ 44.87 \end{bmatrix} + \frac{12}{24}\begin{bmatrix} 50.5 \\ 47.9 \\ 44.6 \end{bmatrix} = \begin{bmatrix} 50.65 \\ 48.275 \\ 44.735 \end{bmatrix}$$

Statistik ujinya adalah :

$$T^2 = 24\overline{X}\begin{bmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix}^t \left(\begin{bmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix} S \begin{bmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix}^t \right) \begin{bmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix} \overline{X}$$

Sehingga didapat nilai T^2 = 110.032.

Kaidah pengambilan keputusannya adalah pembandingan nilai T^2 dengan nilai tabel sebara F dengan db1=2 dan db2=21 pada level 0.05. dapat diketahui bahwa nilai T^2 > nilai F tabel. Dan dapat dikatakan bahwa rataan jumlah kandungan asam askorbat dalam kedua merk sari air jeruk tersebut tidak sama (tidak sejajar dengan sumbu X). Hal ini jelas terlihat pada grafik diatas.

Pengambilan keputusan untuk uji kesamaan dapat juga dilihat berdasarkan nilai statistik *wilks'lamda, pillai's trace, Hottelling-lawleytrace dan Roy's greatest root* yang diperoleh menggunakan *PROC GLM* pada software SAS.

Statistic	Value	F	Num DF	Den DF	Pr > F
Wilks' Lambda	0.12711	17.1683	2	5	0.0058
Pillai's Trace	0.87289	17.1683	2	5	0.0058
Hotelling-Lawley Trace	6.86732	17.1683	2	5	0.0058
Roy's Greatest Root	6.86732	17.1683	2	5	0.0058

Dengan menggunakan α= 0.05, maka dapat disimpulkan bahwa rataan jumlah kandungan asam askorbat dalam kedua merk sari air jeruk tersebut tidak sama (nilai Pr>F lebih kecil dari 0.05).

6.3. Aplikasi SAS

Sebagai ilustrasi dari penggunaan SAS untuk analisis profil telah diuraikan pada ilustrasi sebelumnya. Berikut ini adalah perintah yang digunakan pada SAS dan outputnya.

```
data profil;
input wo w3 w7 merk$;
cards;
52.6  49.4   42.7   Richfood
56.0  48.8   49.2   Sealed
54.2  49.2   48.8   Richfood
48.0  50.3   44.0   Sealed
49.8  42.8   40.4   Richfood
49.6  44.0   42.0   Sealed
46.6  53.2   47.6   Richfood
48.4  48.5   43.2   Sealed
;
proc glm data=profil;
class merk;
model wo w3 w7= merk;
manova h=merk/printe printh;
repeated waktu 3 (0 1 2 );
run;
```

Output dari program di atas sebagai berikut:

```
                              Analisis Profil

                             The GLM Procedure

                          Class Level Information

              Class          Levels     Values

                   merk              2      Richfood Sealed

                   Number of Observations Read           8
                   Number of Observations Used           8
```

```
                                 The GLM Procedure

Dependent Variable: wo

                                        Sum of
Source             DF         Squares     Mean Square    F Value    Pr > F

Model               1      0.18000000      0.18000000       0.01    0.9085
Error               6     75.16000000     12.52666667
Corrected Total     7     75.34000000


                 R-Square     Coeff Var      Root MSE       wo Mean
                 0.002389      6.987765      3.539303      50.65000


Source             DF       Type I SS     Mean Square    F Value    Pr > F
merk                1      0.18000000      0.18000000       0.01    0.9085


Source             DF     Type III SS     Mean Square    F Value    Pr > F
merk                1      0.18000000      0.18000000       0.01    0.9085
```

```
                                 The GLM Procedure

Dependent Variable: w3

                                        Sum of
Source             DF         Squares     Mean Square    F Value    Pr > F
Model               1      1.12500000      1.12500000       0.09    0.7784
Error               6     77.93000000     12.98833333
Corrected Total     7     79.05500000

                 R-Square     Coeff Var      Root MSE       w3 Mean
                 0.014231      7.465423      3.603933      48.27500


Source             DF       Type I SS     Mean Square    F Value    Pr > F
merk                1      1.12500000      1.12500000       0.09    0.7784


Source             DF     Type III SS     Mean Square    F Value    Pr > F
merk                1      1.12500000      1.12500000       0.09    0.7784


Dependent Variable: w7

                                        Sum of
Source             DF         Squares     Mean Square    F Value    Pr > F
Model               1      0.15125000      0.15125000       0.01    0.9175
Error               6     77.82750000     12.97125000
Corrected Total     7     77.97875000

                 R-Square     Coeff Var      Root MSE       w7 Mean
                 0.001940      8.050432      3.601562      44.73750
```

```
Source             DF       Type I SS     Mean Square    F Value    Pr > F
merk                1      0.15125000      0.15125000       0.01    0.9175

Source             DF     Type III SS     Mean Square    F Value    Pr > F
merk                1      0.15125000      0.15125000       0.01    0.9175


                              The GLM Procedure
                      Multivariate Analysis of Variance

                            E = Error SSCP Matrix

                      wo                 w3                 w7

wo             75.16              -8.84              34.54
w3             -8.84              77.93              51.105
w7             34.54              51.105             77.8275


      Partial Correlation Coefficients from the Error SSCP Matrix / Prob > |r|

                 DF = 6             wo             w3             w7

                 wo          1.000000      -0.115507       0.451609
                                              0.8052         0.3091

                 w3         -0.115507       1.000000       0.656213
                               0.8052                        0.1094

                 w7          0.451609       0.656213       1.000000
                           0.3091         0.1094

                              The GLM Procedure
                      Multivariate Analysis of Variance

                      H = Type III SSCP Matrix for merk

                      wo                 w3                 w7

wo              0.18               0.45              0.165
w3              0.45              1.125             0.4125
w7             0.165             0.4125            0.15125


             Characteristic Roots and Vectors of: E Inverse * H, where
                        H = Type III SSCP Matrix for merk
                              E = Error SSCP Matrix

             Characteristic               Characteristic Vector  V'EV=1
           Root    Percent               wo              w3              w7

0.03468049     100.00      0.11814022      0.18137609     -0.14469731
0.00000000       0.00      0.08737238     -0.04939903      0.03940931
0.00000000       0.00     -0.07154281     -0.02602314      0.14901890


  MANOVA Test Criteria and Exact F Statistics for the Hypothesis of No Overall merk
                                     Effect
                        H = Type III SSCP Matrix for merk
                              E = Error SSCP Matrix
```

S=1 M=0.5 N=1

Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.96648193	0.05	3	4	0.9850
Pillai's Trace	0.03351807	0.05	3	4	0.9850
Hotelling-Lawley Trace	0.03468049	0.05	3	4	0.9850
Roy's Greatest Root	0.03468049	0.05	3	4	0.9850

The GLM Procedure
Repeated Measures Analysis of Variance
Repeated Measures Level Information

Dependent Variable	wo	w3	w7
Level of waktu	0	1	2

MANOVA Test Criteria and Exact F Statistics for the Hypothesis of no waktu Effect
H = Type III SSCP Matrix for waktu
E = Error SSCP Matrix
S=1 M=0 N=1.5

Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.12710807	17.17	2	5	0.0058
Pillai's Trace	0.87289193	17.17	2	5	0.0058
Hotelling-Lawley Trace	6.86732130	17.17	2	5	0.0058
Roy's Greatest Root	6.86732130	17.17	2	5	0.0058

MANOVA Test Criteria and Exact F Statistics for the Hypothesis of no waktu*merk Effect
H = Type III SSCP Matrix for waktu*merk
E = Error SSCP Matrix

S=1 M=0 N=1.5

Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.99089826	0.02	2	5	0.9774
Pillai's Trace	0.00910174	0.02	2	5	0.9774
Hotelling-Lawley Trace	0.00918535	0.02	2	5	0.9774
Roy's Greatest Root	0.00918535	0.02	2	5	0.9774

The GLM Procedure
Repeated Measures Analysis of Variance
Tests of Hypotheses for Between Subjects Effects

Source	DF	Type III SS	Mean Square	F Value	Pr > F
merk	1	1.1704167	1.1704167	0.05	0.8227
Error	6	128.1758333	21.3626389		

The GLM Procedure
Repeated Measures Analysis of Variance
Univariate Tests of Hypotheses for Within Subject Effects

						Adj Pr > F	
Source	DF	Type III SS	Mean Square	F Value	Pr > F	G - G	H - F

```
waktu         2   141.6325000  70.8162500      8.27  0.0055    0.0155  0.0063
waktu*merk    2     0.2858333   0.1429167      0.02  0.9835    0.9493  0.9810
Error(waktu) 12    102.7416667 8.5618056

                         Greenhouse-Geisser Epsilon    0.6814
                         Huynh-Feldt Epsilon           0.9599
```

6.4. Latihan

1. Data berikut merupakan nilai skala menurut kelas sosial ekonomi.

	Skala			
Kelas sosial ekonomi	**A**	**B**	**C**	**Total**
1	19	20	18	57
	20	21	19	60
	19	22	22	63
	18	19	21	58
	16	18	20	54
	17	22	19	58
	20	19	20	59
	15	19	19	54
Total	144	160	158	462
2	12	14	12	38
	15	15	17	47
	15	17	15	47
	13	14	14	41
	14	16	13	43
Total	69	76	71	216
3	15	14	17	46
	13	14	15	42
	12	15	15	42
	12	13	13	38

Total	52	56	60	168
4	8	9	10	27
	10	10	12	32
	11	10	10	31
	11	7	12	30
Total	40	36	44	120
Grand Total	305	328	333	966

a. Apakah sejajar profil dari keempat kelas sosial ekonomi tersebut?(gunakan alpha 5 %)
b. Apakah antar kelas sosial ekonomi kondisinya sama ? (Gunakan alpha 5 %)

Petunjuk: Untuk pertanyaan pertama, observasi ditransformasi perbedaan skala A dengan skala B (Skala A-Skala B), dan Skala B-Skala C. Maka Matriks jumlah kuadrat

$$\text{T (Total)} = \begin{bmatrix} 75.81 & -54.48 \\ -54.48 & 85.81 \end{bmatrix}$$

$$\text{P (Perlakuan)} = \begin{bmatrix} 424.61 & -20.48 \\ -20.48 & 24.31 \end{bmatrix}$$

$$\text{E (Galat)} = \begin{bmatrix} 51.20 & -34.00 \\ -34.00 & 61.50 \end{bmatrix}$$

7

7. Analisis Komponen Utama (*Principal Component Analysis*)

7.1. Pendahuluan

Masalah pereduksian dimensi dari segugus data peubah data peubah ganda yang besar sering menjadi topik yang menarik untuk dibahas. Beberapa masalah yang timbul dalam mereduksi dimensi tersebut adalah bagaimana caranya mendapatkan gugus peubah yang lebih kecil namun tetap mampu mempertahankan sebagian besar informasi yang terkandung pada data asal. Salah satu metode yang baik digunakan adalah dengan menggunakan analisis komponen utama. Komponen utama mampu mempertahankan sebagian besar informasi yang diukur menggunakan sedikit peubah yang menjadi komponen utamanya saja. Rao (1964) juga mempertimbangkan ukuran lain untuk menilai informasi yang terkandung sehingga proses pencarian komponen utama dengan batasan informasi yang optimal menjadi beragam.

Analisis peubah ganda merupakan analisis yang sering kali membutuhkan banyak 'sumberdaya' lebih, informasi yang ada pada peubah-peubah penjelasnya (x) seringkali tumpah tindih oleh karena itu kita membutuhkan analisis yang dapat mengatasi masalah ini. Salah satu analisis yang dapat diterapkan untuk mengatasinya adalah analisis komponen utama.

7.2. Komponen Utama

Komponen Utama merupakan kombinasi linear dari peubah yang diamati, informasi yang terkandung pada komponen utama merupakan gabungan dari semua peubah dengan bobot tertentu. Kombinasi linear yang dipilih merupakan kombinasi linear dengan ragam paling besar yang memuat informasi paling banyak . Antar Komponen Utama bersifat ortogonal, tidak berkorelasi dan informasinya tidak tumpang tindih

Hasil dari prosedur ini nantinya digunakan pada analisis lebih lanjut, seperti analisis pengelompokan dan regresi Komponen Utama. Misalkan kita x_1 , … , x_p merupakan p peubah yang menjadi perhatian kita, memiliki sebaran peubah ganda dengan vektor rataan $\boldsymbol{\mu}$ dan matriks peragam $\boldsymbol{\Sigma}$. Komponen Utama seperti telah dijelaskan di atas merupakan kombinasi linier dari p peubah asal, atau dapat ditulis :

$$Y = A X$$

dimana:

$$Y = \begin{bmatrix} Y_1 \\ \dots \\ Y_p \end{bmatrix},\ A = \begin{bmatrix} a_{11} & \dots & a_{1p} \\ \dots & \dots & \dots \\ a_{p1} & \dots & a_{pp} \end{bmatrix},\ X = \begin{bmatrix} X_1 \\ \dots \\ X_p \end{bmatrix}$$

Sehingga komponen utama pertama dapat di tulis sebagai :

$$\mathbf{Y}_1 = a_{11}\,\mathbf{X}_1 + a_{12}\,\mathbf{X}_2 + \dots + a_{1p}\,\mathbf{X}_p$$

$$\mathbf{Y}_1 = \mathbf{a}_1'\,\mathbf{X}$$

Yang memiliki ragam sebesar : $S_{Y1}^2 = var(\boldsymbol{a}_1' X) = \boldsymbol{a}_1'\, S\, \boldsymbol{a}_1$

Pemilihan vektor koefisien Komponen Utama I adalah sedemikian rupa sehingga ragam S_{Y1}^2 terbesar di antara vektor koefisien yang lain. Untuk

mendapatkan hal tersebut dapat dilakukan melalui persamaan Lagrange dengan kendala $\underline{a}_1'\underline{a}_1=1$.

Maksimumkan: $Var(Y_1) = \mathbf{a_1}'\mathbf{x} = \mathbf{a}_1'S\mathbf{a}_1$

dengan kendala : $\mathbf{a}_1'\mathbf{a}_1 = 1$

Persamaan Lagrange :

$$f = (\mathbf{a}_1'S\, \mathbf{a}_1) - \lambda_1 (\mathbf{a}_1'\, \mathbf{a}_1 - 1)$$

untuk mencari titik kritis dari persamaan langrange tersebut dapat dilakukan dengan cara mencari turunan pertama sebagai berikut:

$$df/d\mathbf{a_1} = 2\, S\mathbf{a}_1 - 2\, \lambda_1\mathbf{a}_1 = 0$$

$S\mathbf{a}_1 = \lambda_1\mathbf{a}_1$ (ini merupakan persamaan ciri dari matriks S)

Jika kedua ruas dikalikan dengan $\mathbf{a}_1'$ maka ruas kiri merupakan ragam dari Y1,

$$\mathbf{a}_1'S\mathbf{a}_1 = \mathbf{a}_1'\lambda_1\mathbf{a}_1 = \lambda_1 = Var(Y_1)$$

Dengan demikian agar $Var(Y_1)$ maksimum maka haruslah λ_1 merupakan akar ciri terbesar dari matriks peragam S dan $\mathbf{a}_1$ merupakan vektor ciri yang berpadanan dengan λ_1.

Selanjutnya untuk Komponen Utama kedua dapat ditulis sebagai :

$\mathbf{Y}_2 = a_{21}\, \mathbf{X}_1 + a_{22}\, \mathbf{X}_2 + \ldots. + a_{2p}\, \mathbf{X}_p$

$\mathbf{Y}_2 = \mathbf{a}_2'\, \mathbf{X}$

Yang memiliki ragam sebesar : $S_{Y2}^2 = var(\boldsymbol{a}_2'X) = \boldsymbol{a}_2'\, S\, \boldsymbol{a}_2$

Pemilihan vektor koefisien $\mathbf{a}_2$ adalah sedemikian rupa sehingga ragam S^2_{Y2} maksimum dengan kendala $\mathbf{a}_2'\mathbf{a}_2 = 1$ dan Cov(Y_1,Y_2)=$\mathbf{a}_1'S\mathbf{a}_2 = 0$ atau $\mathbf{a}_1'\mathbf{a}_2 = 0$. Untuk mendapatkan hal tersebut, prosedur yang digunakan mirip seperti pemilihan vektor koefisien untuk Komponen Utama I, yaitu dengan persamaan Lagrange.

Maksimumkan: Var(Y_2) = $\mathbf{a_2'x} = \mathbf{a}_2'S\mathbf{a}_2$

dengan kendala : $\mathbf{a}_2'\mathbf{a}_2 = 1$ dan $\mathbf{a}_1'\mathbf{a}_2 = 0$

Persamaan Lagrange :

$$f = (\mathbf{a}_2'S\,\mathbf{a}_2) - \lambda_2(\mathbf{a}_2'\,\mathbf{a}_2 - 1) - \alpha\,(\mathbf{a}_1'\,\mathbf{a}_2 - 0)$$

$$= (\mathbf{a}_2'S\,\mathbf{a}_2) - \lambda_2(\mathbf{a}_2'\,\mathbf{a}_2 - 1) - \alpha\,\mathbf{a}_1'\,\mathbf{a}_2$$

Turunan pertama terhadap $\mathbf{a}_2$ diperoleh sebagai berikut:

$$\frac{\partial f}{\partial \boldsymbol{a}2} = 2S\boldsymbol{a}_2 - 2\lambda_2\boldsymbol{a}_2 - \alpha\boldsymbol{a}_1 = 0$$

Jika persamaan tersebut dikalikan dengan $\mathbf{a}_1'$ didapatkan :

$$2\boldsymbol{a}_1'S\boldsymbol{a}_2 - 2\lambda_2\boldsymbol{a}_1'\boldsymbol{a}_2 - \alpha\boldsymbol{a}_1'\boldsymbol{a}_1 = 0$$

Karena $\mathbf{a}_1'\mathbf{a}_1 = 1$ dan $\mathbf{a}_1'\mathbf{a}_2 = 0$ serta $\mathbf{a}_1'S\mathbf{a}_2 = 0$ maka haruslah α=0. Dengan demikian turunan pertama persamaan langrange dapat ditulis menjadi:

$$2S\boldsymbol{a}_2 - 2\lambda_2\boldsymbol{a}_2 - 0\boldsymbol{a}_1 = 2S\boldsymbol{a}_2 - 2\lambda_2\boldsymbol{a}_2 = S\boldsymbol{a}_2 - \lambda_2\boldsymbol{a}_2 = 0$$

$$S\boldsymbol{a}_2 = \lambda_2\boldsymbol{a}_2$$

Ini juga merupakan persamaan ciri dari matriks S. Jika kedua ruas dikalikan dengan $\mathbf{a}_2$' maka sisi kiri merupakan ragam dari $\mathbf{Y}_2$:

$$\mathbf{a}_2'S\mathbf{a}_2 = \mathbf{a}_2'\lambda_2\mathbf{a}_2 = \lambda_2 = Var(Y_2)$$

Agar var(Y_2) maksimum maka haruslah λ_2 merupakan akar ciri terbesar dari S. Karena akar cirri terbesar pertama merupakan ragam dari komponen utama I maka λ_2 dipilih akar ciri terbesar kedua dari matriks S. Demikian juga dengan vector $\mathbf{a}_2$ merupakan vektor ciri yang berpadanan dengan akar ciri terbesar kedua λ_2.

Selanjutnya untuk Komponen Utama ke-i ; i= 3, 4. …, p didapatkan dari kombinasi linier p peubah asal yang memaksimumkan var($\mathbf{a}_i'\mathbf{X}$) dengan kendala $\mathbf{a}_i'\mathbf{a}_i = 1$ dan cov ($\mathbf{a}_i'\mathbf{X}, \mathbf{a}_k'\mathbf{X}$) = 0 untuk k < i.

Seperti telah dijelaskan di atas bahwa :

$$Var(Y_1) = \boldsymbol{a}_1'S\boldsymbol{a}_1 = \lambda_1$$

$$Var(Y_2) = \boldsymbol{a}_2'S\boldsymbol{a}_2 = \lambda_2$$

Dimana λ_1 dan λ_2 merupakan akar cirri terbesar pertama dan kedua dari matriks S. Dengan cara yang sama seperti penurunan pada komponen utama I dan komponen utama II akan diperoleh hasil secara umum sebagai berikut:

$$Var(Y_i) = \boldsymbol{a}_i'S\boldsymbol{a}_i = \lambda_i \text{ untuk i=3, 4, 5, …, p}$$

Dimana λ_i merupakan akar ciri terbesar ke-i dari matriks S dan vector ai merupakan vector ciri yang berpadanan dengan λ_i.

Selanjutnya akan dijelaskan besarnya kontribusi keragaman masing-masing komponen utama dalam menjelaskan keragaman data asal. Untuk menjelaskan hal ini kita harus melihat kembali salah satu teorema dalam aljabar matriks yaitu bahwa teras suatu matriks akan sama dengan jumlah dari akar ciri-akar ciri matriks tersebut. Sehingga dalam kasus ini dapat dituliskan sebagai berikut:

$$teras(S) = \sum_{i=1}^{p} \lambda_i$$

$$s_{11} + s_{22} + \cdots + s_{pp} = \lambda_1 + \lambda_2 + \cdots + \lambda_p$$

$$var(X_1) + var(X_2) + \cdots + var(X_p) = var(Y_1) + var(Y_2) + \cdots + var(Y_p)$$

Berdasarkan hubungan ini maka besarnya kontribusi keragaman relative yang mampu dijelaskan oleh komponen utama ke-I adalah sebesar:

$$\frac{var(Y_i)}{total\ keragaman\ dari\ peubah\ asal} = \frac{\lambda_i}{s_{11}+s_{22}+\cdots+s_{pp}} = \frac{\lambda_i}{\lambda_1+\lambda_2+\cdots+\lambda_p}$$

Atau dapat juga dinyatakan dalam besaran persen sebagai berikut:

$$\frac{\lambda_i}{\sum_{i=1}^{p} \lambda_i}\ x\ 100\%$$

Sedangkan besarnya keragaman kumulatif untuk q buah komponen utama dapat dituliskan sebagai berikut:

$$\frac{\sum_{j=1}^{q} \lambda_j}{\sum_{i=1}^{p} \lambda_i}\ x\ 100\%$$

Apabila kita menghadapi masalah perbedaan skala pengukuran, maka peubah dibakukan menjadi:

$$Z_{ij} = \frac{X_{ij} - \overline{X}_j}{S_j}$$

7.3. Penentuan penggunaan matriks korelasi dan Ragam peragam

Secara umum cukup sulit untuk menentukan kapan menggunakan matriks korelasi atau matriks ragam peragam untuk memperoleh komponen utama. Kesulitan terjadi karena tidak ada hubungan yang jelas antara akar ciri dan vektor ciri matriks ragam peragam dengan matriks korelasi, dan komponen utama yang dihasilkan oleh keduanya dapat sangat berbeda. Demikian juga dengan berapa banyak komponen utama yang digunakan.

Penggunaan matriks korelasi memang cukup efektif kecuali pada dua hal.

Pertama, secara teori pengujian statistik terhadap akar ciri dan vektor ciri matriks korelasi jauh lebih rumit.

Kedua, dengan menggunakan matriks korelasi kita memaksakan setiap peubah memiliki ragam yang sama sehingga tujuan mendapatkan peubah yang kontribusinya paling besar tidak tercapai.

7.4. Penentuan Banyaknya Komponen Utama

Terdapat tiga metode yang umum digunakan untuk menentukan banyaknya komponen utama, yaitu:

❖ <u>**Metode 1**</u>

- didasarkan pada kumulatif proporsi keragaman total yang mampu dijelaskan.

- Metode ini merupakan metode yang paling banyak digunakan, dan dapat diterapkan pada penggunaan matriks korelasi maupun matriks ragam peragam.
- Minimum persentase kergaman yang mampu dijelaskan ditentukan terlebih dahulu, dan selanjutnya banyaknya

komponen yang paling kecil hingga batas itu terpenuhi dijadikan sebagai banyaknya komponen utama yang digunakan.

- Tidak ada patokan baku berapa batas minimum tersebut, sebagian bukau menyebutkan 70%, 80%, bahkan ada yang 90%.

- **Metode 2**
  - hanya dapat diterapkan pada penggunaan matriks korelasi. Ketika menggunakan matriks ini, peubah asal ditransformasi menjadi peubah yang memiliki ragam sama yaitu satu.
  - Pemilihan komponen utama didasarkan pada ragam komponen utama, yang tidak lain adalah akar ciri. Metode ini disarankan oleh Kaiser (1960) yang berargumen bahwa jika peubah asal saling bebas maka komponen utama tidak lain adalah peubah asal, dan setiap komponen utama akan memiliki ragam satu.
  - Dengan cara ini, komponen yang berpadanan dengan akar ciri kurang dari satu tidak digunakan. Jollife (1972) setelah melakukan studi mengatakan bahwa cut off yang lebih baik adalah 0.7.

- **Metode 3**
  - penggunaan grafik yang disebut plot scree.
  - Cara ini dapat digunakan ketika titik awalnya matriks korelasi maupun ragam peragam.
  - Plot scree merupakan plot antara akar ciri λk dengan k.
  - Dengan menggunakan metode ini, banyaknya komponen utama yang dipilih, yaitu k, adalah jika pada titik k tersebut plotnya curam ke kiri tapi tidak curam di kanan. Ide yang ada di belakang metode ini adalah bahwa banyaknya komponen utama yang dipilih sedemikian rupa sehingga selisih antara akar ciri yang berurutan sudah tidak besar lagi. Interpretasi terhadap plot ini sangat subjektif.

7.5. Manfaat lain dari komponen utama

Dalam beberapa kasus terapan, analisis komponen utama dimanfaatkan untuk berbagai keperluan analisis seperti eksplorasi posisi objek dan penangan masalah kolinear antar peubah. Eksplorasi posisi objek diperlukan sebagai alat bantu dalam analisis gerombol. Eksplorasi ini dapat dilakukan dengan membuat plot skor komponen utama pertama dengan kedua. Dari plot ini diharapkan dapat terlihat banyaknya kumpulan objek yang terbentuk, yang pada nantinya dapat digunakan sebagai nilai awal dalam penentuan jumlah gerombol yang akan dibangun.

Sedangkan dalam kasus kolinear antar peubah, yang merupakan salah satu masalah yang sering dihadapi dalam analisis regresi, komponen utama dapat merupakan salah satu solusi dalam mengatasi masalah kolinear. Penerapan ini relevan dengan sifat dari komponen utama yang dibangun yaitu antar komponen utama bersifat saling orthogonal atau saling bebas. Dengan memanfaatkan komponen utama ini masalah kolinear dalam analisis regresi dapat diatasi. Sifat ini juga diperlukan dalam penerapan jarak Euclid, dimana konsep jarak ini juga memerlukan kebebasan antar peubah.

7.6. Penerapan Analisis Komponen Utama dalam Analisis Regresi

Berikut ini disajikan ilustrasi penggunaan Analisis Komponen Utama dalam mengatasi masalah multikolinieritas dalam analisis regresi berganda. Ilustrasi ini diambil dari skripsi Khristiningrum (1997).

Tahap-tahap yang dilakukan dalam analisis regresi Komponen Utama adalah :

1. Peubah bebas asal X_j dibakukan dan diskalakan.

$$Z_{jj} = \left(X_{ji} - \overline{X}_j\right) / S_{jj}^{1/2}$$

2. Menghitung nilai akar ciri (λ_j), vektor ciri (γ_j) dan skor komponen utama (W_j).
3. Meregresikan peubah respon (Y) dengan skor komponen utama (W_j) yang terpilih.

$$Y = \delta_0 + \delta_1 W_1 + \delta_2 W_2 + \ldots + \delta_p W_p$$

4. Mentransformasi persamaan regresi dengan peubah bebas W_j ke peubah bebas Z_j.

$$Y = \beta_0 + \beta_1 Z_1 + \beta_2 Z_2 + \ldots + \beta_r Z_r$$

dengan

$\beta 0 = \delta 0$

$\beta j = \gamma j1\ \delta 1 + \gamma j2\ \delta 2 + \ldots + \gamma jp\ \delta p$; j = 1, 2, …, r

5. Mentransformasi persamaan regresi dengan peubah bebas Z_j ke peubah bebas X_j.

$$Y = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \ldots + \alpha_r X_r$$

Ragam koefisien regresi peubah bebas X adalah :

$$s^{*2} = \frac{s^2}{\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2}$$

sedangkan

$$\sigma_{\alpha j}^2 = s^{*2} \sum_{i=1}^{p} \frac{\gamma_{ji}^2}{\lambda_i}$$

s^2 adalah ragam galat dari persamaan pada point (3).

Pengujian keberartian koefisien regresi, dilakukan dengan menggunakan uji t-student dengan statistik uji :

$$t(\alpha_j) = \frac{\alpha_j}{\sqrt{\sigma_{\alpha j}^2}}$$

7.7. Aplikasi SAS

Prosedur **PRINCOMP** merupakan prosedur yang biasa mengerjakan analisis komponen utama. Bentuk umum dari perintah **PROC PRINCOMP** adalah :

PROC PRINCOMP <*options*>;

VAR *variables*;

RUN;

Berikut ini ilustrasi penggunaan SAS untuk analisis komponen utama.

Kejahatan merupakan suatu hal yang tidak dapat diabaikan dalam pengelolaan sebuah wilayah. Dari setiap negara bagian di Amerika Serikat, dicatat besarnya beberapa macam kejahatan selama satu tahun dalam bentuk indeks kejahatan per 100.000 penduduk. Dari data ini ingin diketahui pola dari data kejahatan. Kejahatan-kejahatan yang dicatat adalah :

- Murder (Pembunhuan)
- Rape
- Robbery (perampokan)
- Assault
- Burglary
- Larceny
- Auto_Theft (Pencurian Kendaraan Bermotor)

Pertama adalah kita harus memasukkan data, secara in-stream bisa dituliskan sebagai berikut :

```
Data Crime;
   title 'Crime Rates per 100,000 Population by State';
   input State $1-15 Murder Rape Robbery Assault
Burglary Larceny Auto_Theft;
   datalines;
Alabama        14.2 25.2  96.8 278.3 1135.5 1881.9 280.7
Alaska         10.8 51.6  96.8 284.0 1331.7 3369.8 753.3
Arizona         9.5 34.2 138.2 312.3 2346.1 4467.4 439.5
Arkansas        8.8 27.6  83.2 203.4  972.6 1862.1 183.4
California     11.5 49.4 287.0 358.0 2139.4 3499.8 663.5
Colorado        6.3 42.0 170.7 292.9 1935.2 3903.2 477.1
..............
;
```

catatan : Dalam analisis yang sesungguhnya terhadap sebuah data, perlu dilakukan eksplorasi secara lengkap, termasuk pembuatan plot dan grafik, analisis korelasi dan statistika deskriptif. Hanya eksplorasi singkat saja yang ditunjukkan dalam model ini. Sebagai bagian dari kegiatan eksplorasi, kita akan lihat korelasi antara peubah yang terlibat.

```
proc corr data = crime nosimple noprob;
var murder--auto_theft;
title 'Korelasi antar peubah kejahatan';
run;
```

Keluaran dari perintah di atas adalah

```
                        Korelasi antar peubah kejahatan
                              The CORR Procedure
   7 Variables: Murder Rape    Robbery     Assault     Burglary    Larceny
                                  Auto_Theft
                    Pearson Correlation Coefficients, N = 50

                                                                      Auto_
              Murder    Rape    Robbery    Assault  Burglary  Larceny   Theft

Murder       1.00000   0.60122 0.48371   0.64855    0.38582  0.10192  0.06881
Rape         0.60122   1.00000 0.59188   0.74026    0.71213  0.61399  0.34890
Robbery      0.48371   0.59188 1.00000   0.55708    0.63724  0.44674  0.59068
Assault      0.64855   0.74026 0.55708   1.00000    0.62291  0.40436  0.27584
Burglary     0.38582   0.71213 0.63724   0.62291    1.00000  0.79212  0.55795
Larceny      0.10192   0.61399 0.44674   0.40436    0.79212  1.00000  0.44418
Auto_Theft   0.06881   0.34890 0.59068   0.27584    0.55795  0.44418  1.00000
```

Jika kita amati secara seksama tabel korelasi di atas, kita lihat bahwa beberapa peubah saling berkorelasi. Misalnya saja burglary berkorelasi dengan larceny, dengan koefisien korelasi sebesar 0.79212.

Pada contoh di atas, untuk mengeluarkan semua komponen utama yaitu sebanyak peubah yang terlibat, perintah yang diberikan adalah sebagai berikut :

```
proc princomp data=crime;
  var murder--auto_theft;
  title 'Komponen Utama untuk Data Kejahatan';
run;
```

Output yang dihasilkan terdiri atas beberapa bagian yang pelu untuk diketahui. Pertama, PROC PRINCOMP membuat daftar banyaknya peubah dan objek yang diamati pada gugus data. (juga ditampilkan statistik deskriptif dan korelasi antar peubah, tapi tidak ditampilkan di sini. Penggalan dari keluaran PROC PRINCOMP tersebut adalah :

```
           Komponen Utama untuk Data Kejahatan
                 The PRINCOMP Procedure
                Observations          50
                Variables              7
```

Selanjutnya, PROC PRINCOMP menampilkan akar ciri dan vector ciri dari matriks korelasi. Ingat bahwa akar ciri diinterpretasikan sebagai ragam dari kombinsai-kombinasi linear atau komponen, dan vektor ciri sebagai bobot yang digunakan untuk menyusun skor komponen utama dari data yang terpusat, yaitu nilai asli dikurangi rataan peubahnya. Penggalan dari keluaran PROC PRINCOMP mengenai hal ini adalah :

Eigenvalues of the Correlation Matrix

	Eigenvalue	Difference	Proportion	Cumulative
1	4.11495951	2.87623768	0.5879	0.5879
2	1.23872183	0.51290521	0.1770	0.7648
3	0.72581663	0.40938458	0.1037	0.8685
4	0.31643205	0.05845759	0.0452	0.9137
5	0.25797446	0.03593499	0.0369	0.9506
6	0.22203947	0.09798342	0.0317	0.9823
7	0.12405606		0.0177	1.0000

Komponen utama pertama, mengandung hampir 60 persen keragaman data asal, dan tiga komponen pertama secara bersama mampu menjelaskan 87 persen keragaman data. Masing-masing komponen sisanya hanya memiliki kontribusi keragaman kurang dari lima persen. Hal ini menyarankan bahwa kita bisa membuang komponen-komponen akhir, hanya mengambil tiga komponen utama pertama, tanpa kehilangan banyak informasi.

Vektor-vektor ciri bisa diinterpretasikan sebagai koefisien yang menyusun skor komponen dari data terpusat. Kadangkala, kita membakukan vector ciri sehingga panjang vektornya sama dengan akar dari akar ciri yang bersesuaian. Namun ini bukan keluaran default dari PROC PRINCOMP. Selanjutnya kita bahas hanya menggunakan tiga komponen utama pertama, karena komponen sisanya hanya menerangkan sebagian kecil informasi. Penggalan selanjutnya dari keluaran PROC PRINCOMP adalah sebagai berikut :

```
                   The PRINCOMP Procedure

                        Eigenvectors

              Prin1        Prin2        Prin3
Murder        0.300279     -.629174     0.178245
Rape          0.431759     -.169435     -.244198
Robbery       0.396875     0.042247     0.495861
Assault       0.396652     -.343528     -.069510
Burglary      0.440157     0.203341     -.209895
Larceny       0.357360     0.402319     -.539231
Auto_Theft    0.295177     0.502421     0.568384
```

Jika kita mengamati besarnya koefisien dari KU 1 terlihat bahwa setiap koefisien memiliki nilai positif dan hampir sama besar. Kondisi ini merupakan hal yang umum terjadi dalam analisis komponen utama karena KU 1 umumnya merupakan rata-rata dari semua peubah. Koefisien ini tidak bisa diartikan sebagai korelasi antara komponen dengan peubah asli, kalau ingin diartikan seperti itu harus dilakukan penskalaan, dengan membuat panjang vector sama dengan akar dari akar ciri yang berpadanan.

Sementara itu, KU 2 memiliki nilai koefisien yang negatif di beberapa peubah dan positif pada peubah lain. Komponen ini merupakan kontras antara kejahatan pencurian dengan tindak kekerasan. Sedangkan KU 3 merupakan kontras antara pencurian tanpa dan dengan kekerasan.

Seringkali interpretasi komponen merupakan hal yang sulit dan tidak mungkin. Dalam kasus ini kita gunakan analisis komponen utama sebagai upaya matematis untuk menyederhanakan banyaknya peubah. Jika kita ingin menginterpretasikan komponen sebagai peubah laten apa yang berada di belakang peubah yang diamati, akan lebih tepat jika digunakan analisis faktor (factor analysis).

Penggunaan AKU untuk Matriks Korelasi

Kadang kala kita tidak lagi memiliki matrik data yang berukuran n x p, namun kita memiliki matirks korelasi antar peubah asal. PROC PRINCOMP bisa melakukan analisis komponen utama dengan masukan berupa matriks korelasi dengan cara sesederhana kalau kita gunakan matriks data, seperti contoh sebelumnya.

Misalkan dari data yang sama dengan sebelumnya, kita buat terlebih dahulu matriks korelasi. Perintah yang diberikan adalah :

```
proc corr data = crime noprint out=cri_corr;
var murder--auto_theft;
       title 'Korelasi antar peubah kejahatan';
run;
```

Perintah di atas menghasilkan sebuah gugus data baru yang bernama cri_corr. Untuk melihat isi dari gugus data ini, kita bisa mencetak menggunakan perintah :

```
proc print data = cri_corr;
   title ‘Gugus data hasil PROC CORR’;
run;
```

Hasilnya adalah sebagai berikut :

```
                       Gugus data hasil PROC CORR    2




                                                                          Auto_
Obs  _TYPE_  _NAME_  Murder  Rape  Robbery Assault Burglary Larceny Theft

1     MEAN          7.4440 25.7340 124.092 211.300 1291.90 2671.29 377.526
2     STD           3.8668 10.7596  88.349 100.253  432.46  725.91 193.394
3     N            50.0000 50.0000  50.000  50.000   50.00   50.00  50.000
4     CORR  Murder 1.0000  0.6012  0.484    0.649    0.39    0.10   0.069
5     CORR  Rape   0.6012  1.0000  0.592    0.740    0.71    0.61   0.349
6     CORR  Robbery0.4837  0.5919  1.000    0.557    0.64    0.45   0.591
7     CORR  Assa   0.6486  0.7403  0.557    1.000    0.62    0.40   0.276
8     CORR  Burglary0.3858 0.7121  0.637    0.623    1.00    0.79   0.558
9     CORR  Larceny 0.1019 0.6140  0.447    0.404    0.79    1.00   0.444
10    CORR  Auto_Theft0.0688 0.3489  0.591  0.276    0.56    0.44
1.00
```

Selanjutnya kita lakukan analisis komponen utama, menggunakan PROC PRINCOMP.

```
proc princomp data=cri_corr (type = corr);
  title 'Hasil AKU dengan input matriks korelasi';
run;
```

```
              Hasil AKU dengan input matriks korelasi
                      The PRINCOMP Procedure
                     Observations          50
                     Variables              7
              Eigenvalues of the Correlation Matrix
     Eigenvalue    Difference    Proportion    Cumulative
1    4.11495951    2.87623768        0.5879        0.5879
2    1.23872183    0.51290521        0.1770        0.7648
3    0.72581663    0.40938458        0.1037        0.8685
4    0.31643205    0.05845759        0.0452        0.9137
5    0.25797446    0.03593499        0.0369        0.9506
6    0.22203947    0.09798342        0.0317        0.9823
7    0.12405606                      0.0177        1.0000
                          Eigenvectors
             Prin1    Prin2    Prin3    Prin4    Prin5     Prin6     Prin7
Murder    0.300279 -.629174 0.178245 -.232114 0.538123  0.259117  0.267593
Rape      0.431759 -.169435 -.244198 0.062216 0.188471  -.773271  -.296485
Robbery   0.396875 0.042247 0.495861 -.557989 -.519977  -.114385  -.003903
Assault   0.396652 -.343528 -.069510 0.629804 -.506651  0.172363  0.191745
Burglar   0.440157 0.203341 -.209895 -.057555 0.101033  0.535987  -.648117
Larceny   0.357360 0.402319 -.539231 -.234890 0.030099  0.039406  0.601690
Auto_Theft0.295177 0.5024210.568384 0.419238 0.369753  -.057298  0.147046
```

7.8. Latihan

1. Diberikan matriks kovarian sebagai berikut:

$$\sum = \begin{bmatrix} 6 & 3 \\ 3 & 4 \end{bmatrix}$$

   a. Buatlah komponen utama Y_1 dan Y_2 berdasarkan matriks kovarian tersebut
   b. Hitung proporsi keragaman total yang dapat dijelaskan oleh komponen utama1!

2. Konversi matriks kovarian pada soal no 1 ke dalam matriks korelasi ρ. Berdasarkan matriks korelasi,
   a. Hitung nilai komponen utamanya
   b. Hitung proporsi keragaman total yang dapat dijelaskan oleh komponen utama 1!

c. Bandingkan hasil yang diperoleh dengan hasil pada no 1, apakah hasilnya sama? Bagaimana seharusnya?

3. Diberikan matriks kovarian sebagai berikut:

$$\sum = \begin{bmatrix} 4 & 0 & 0 \\ 0 & 8 & 0 \\ 0 & 0 & 4 \end{bmatrix}$$

a. Tentukan komponen utama Y1 , Y2 dan Y3.

b. Apa yang dapat Anda ceritakan mengenai komponen utama yang dihasilkan!

c. Berdasarkan matriks kovarian diatas, tunjukkanlah bahwa ragam dari komponen utama pertama sama dengan akar ciri terbesar dari matriks kovariannya

4. Seorang peneliti melaporkan matriks korelasi dari hasil pengukuran efikasi diameter, susut bobot, respirasi O_2 dan total padatan terlarut

$$\begin{bmatrix} 1 & 0.81 & -0.82 & -0.8 \\ 0.81 & 1 & -0.91 & -0.8 \\ -0.82 & -0.91 & 1 & 0.86 \\ -0.8 & -0.8 & 0.86 & 1 \end{bmatrix}$$

a. Hitunglah komponen utama pertama dari matrik korelasi tersebut

b. Hitung proporsi keragaman total yang dapat dijelaskan oleh komponenutama 1

5. Seorang peneliti tanaman melakukan penelitian mengenai karakteristik kuantitatif generatif dan vegetatif bunga dari suatu tanaman. Peubah yang diamati adalah:

- bobot kering biji (X1) - jumlah petal (X2)
- jumlah sepal (X3) - jumlah cabang produktif (X4)
- waktu mekar (X5) - dimater batang (X6)
- dimaeter cabang (X7) - jumlah buku pembibitan (X8)

Data tersebut diolah dengan menggunakan analisis komponen utama:

```
Principal Component Analysis: x1; x2; x3; x4; x5; x6; x7; x8

Eigenanalysis of the Correlation Matrix
```

Eigenvalue	6,6220	0,5726	0,3409	0,3213	0,0715	0,0574	0,0145	0,0000
Proportion	0,828	0,072	0,043	0,040	0,009	0,007	0,002	0,000
Cumulative	0,828	0,899	0,942	0,982	0,991	0,998	1,000	1,000

Variable	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
x1	0,366	-0,324	-0,216	-0,104	-0,154	-0,736	-0,373	0,000
x2	0,307	-0,783	0,147	0,007	-0,244	0,443	0,122	-0,000
x3	0,338	0,327	-0,685	-0,051	-0,449	0,324	0,003	-0,000
x4	0,373	-0,061	-0,271	0,028	0,817	0,239	-0,240	0,000
x5	0,383	0,069	-0,006	-0,143	0,165	-0,283	0,849	0,000
x6	0,359	0,206	0,285	0,545	-0,089	-0,017	-0,073	-0,661
x7	0,329	0,273	0,454	-0,723	-0,064	0,142	-0,230	-0,107
x8	0,367	0,222	0,319	0,381	-0,088	0,005	-0,098	0,743

Pertanyaanya :

a. Menurut Anda, Apa alasan peneliti tersebut menggunakan analisis komponen utama?
b. Pada kasus dan kondisi data yang bagaimanakah analisis komponen utama dilakukan?
c. Berdasarkan hasil diatas, ada berapa komponen utama yang sebaiknya dipilih? Berikan alasan Anda!

8

8.Analisis Faktor (*Factor Analysis*)

8.1. Pendahuluan

Analisis faktor adalah salah satu analisis yang banyak digunakan pada statistik peubah ganda, diperkenalkan pertama kali oleh Spearman (1904), dan dikembangkan oleh Thurstone (1947), Thomson (1951), Lawley (1940, 1941) dan lainnya. Pada awalnya analisis ini tergolong sulit dan kontroversial, namun dalam perkembangaannya dirasakan menjadi alat yang sangat berguna. Terutama setelah perkembangan komputer dan paket-paket piranti lunak statistik., serta pada tahun 1970-an banyak terbit buku dan publikasi lain yang mambahas penerapannya di berbagai bidang seprti biologi, kimia, ekologi, ekonomi, pendidikan, ilmu politik, psikologi, dan sosiologi. Penerapan tentang analisis ini secara detail di berbagai bidang dapat dirujuk pada berbagai terbitan.

Pada analisis faktor, perhatian utama adalah menemukan hubungan internal antar segugus peubah acak. Spearman (1904) ketika berkerja dengan skor-skor yang diperoleh dari pengujian pendidikan, mencatat pola sistematik pada matriks korelasi antar skor. Ilustrasinya adalah sebagai berikut, misalkan matriks korelasi ρ dihasilkan dari enam buah skor/nilai hasil pengujian : Sejarah (C), Bahasa Perancis (F), Bahasa Inggris (E), Matematika (M), Ketrampilan Tangan (D), dan Music (Mu), yang diukur pada 33 siswa sekolah persiapan. Matriks korelasi yang dihasilkan adalah

$$\rho = \begin{matrix} C \\ F \\ E \\ M \\ D \\ Mu \end{matrix} \begin{bmatrix} 1.00 & 0.83 & 0.78 & 0.70 & 0.66 & 0.63 \\ 0.83 & 1.00 & 0.67 & 0.67 & 0.65 & 0.57 \\ 0.78 & 0.67 & 1.00 & 0.64 & 0.54 & 0.51 \\ 0.70 & 0.67 & 0.64 & 1.00 & 0.45 & 0.51 \\ 0.66 & 0.65 & 0.54 & 0.45 & 1.00 & 0.40 \\ 0.63 & 0.57 & 0.51 & 0.51 & 0.40 & 1.00 \end{bmatrix}$$

Peubah-peubah di atas disusun sedemikian rupa sehingga didapatkan pola korelasi yang menurun pada matriks tersebut. Yang menarik pada matriks ini tentus saja bukanlah unsur diagonalnya. Jika diperhatikan sesecara seksama, terlihat bahwa setiap dua baris pada matriks itu hampir proporsional. Misalnya saja perhatikan baris untuk C dan M, rasionya adalah

$$\frac{0.83}{0.67} \approx \frac{0.78}{0.64} \approx \frac{0.66}{0.45} \approx \frac{0.63}{0.51} \approx 1.2$$

Spearman berpendapat bahwa pola pada matriks korelasi itu dapat dijelaskan menggunakan model

$$x_i = L_i f + \varepsilon_i ; \quad i = 1, \ldots, p \ (p = 6) \qquad (1)$$

dengan x_i mewakili pengamatan pada peubah ke-*i*, *f* adalah peubah acak yang menerangkan faktor umum (diidentifikasi sebagai tingkat intelegensia) bersama untuk semua peubah x_i, dan ε_i adalah peubah acak yang mewakili faktor khusus (diidentifikasi sebagai kemampuan khusus pada uji pelajaran tertentu) yang hanya ada pada masing-masing x_i, dan L_i adalah konstanta yang tak diketahui. Dengan mengasumsikan bahwa $var(f) = \sigma_f^2$, $var(\varepsilon_i) = \sigma_i^2$, $cov(\varepsilon_i, \varepsilon_j) = 0$, untuk $i \neq j$, dan $cov(\varepsilon_i, f) = 0$, rasio dari koefisien korelasi akan memenuhi sifat konstan. Jelasnya, untuk $i \neq j \neq k$, $i = 1, \ldots, p$,

$$\frac{\operatorname{cov}(x_i, x_j)}{\operatorname{cov}(x_i, x_k)} = \frac{L_j}{L_k}$$

tidak tergantung pada *i*. Spearman menggunakan pengamatan ini untuk menguatkan pendapatnya bahwa ‘semua cabang kegiatan intelektual memiliki fungsi dasar yang sama, sedangkan unsur spesifik di setiap kegiatan berbeda-beda’. Dua faktor di atas (faktor bersama dan faktor khusus) kedengaran kurang memadai untuk menerangkan fenomena tersebut. Sehingga, secara alamiah beberapa faktor bersama dan satu faktor khusus dimasukkan dalam model yang dikenal dengan model faktor.

8.2. Model Faktor

Misalkan **x** adalah vektor acak dengan vektor rata-rata $\boldsymbol{\mu}$ dan matriks ragam peragam $\boldsymbol{\Sigma}$, dan hubungan antar unsur vektor **x** dapat dituliskan dalam model faktor

$$\mathbf{x} = \boldsymbol{\mu} + \mathbf{Lf} + \boldsymbol{\varepsilon} \qquad (2)$$

dimana $\boldsymbol{\mu}$ adalah vektor konstanta, **f** adalah vektor acak dengan ukuran *kx1* ($k < p$), dengan unsur $f_1, \ldots, f_k$, dan disebut faktor bersama, **L** adalah matriks konstanta yang tidak diketahui nilainya berukuran pxk, disebut loading faktor, dan unsur-unsur $\varepsilon_1, \ldots, \varepsilon_p$ adalah unsur vektor acak $\boldsymbol{\varepsilon}$ yang disebut faktor khusus. Dalam hal ini diasumsikan bahwa vektor **f** dan $\boldsymbol{\varepsilon}$ saling tidak berkorelasi. Jadi model di atas berimplikasi bahwa untuk unsur **x** tertentu, misalkan x_i yang mewakili pengukuranpada peubah tertentu dapat dituliskan sebagai kombinasi linear dari seluruh faktor bersama dan sebuah faktor khusus ε_i. Atau dituliskan,

$$\begin{aligned} x_1 &= \mu_1 + l_{11}f_1 + \ldots + l_{1k}f_k + \varepsilon_1 \\ x_2 &= \mu_2 + l_{21}f_1 + \ldots + l_{2k}f_k + \varepsilon_2 \\ &\cdot \\ &\cdot \\ &\cdot \\ x_p &= \mu_p + l_{p1}f_1 + \ldots + l_{pk}f_k + \varepsilon_p \end{aligned}$$

dimana l_{ij}, unsur ke-(i, j) dari matriks **L**, adalah loading faktor untuk faktor bersama f_j terhadap x_i. Jika k = 1 maka model faktor tersebut tereduksi menjadi model Spearman seperti pada sub bab sebelumnya.

Dalam permasalahan analisis faktor yang dilakukan adalah menentukan faktor-faktor bersama sehingga korelasi di antara unsur vektor **x** terangkum pada faktor-faktor yang diperoleh.

Model faktor di atas menggunakan asumsi bahwa

- $E(\mathbf{f}) = \mathbf{0}$
- $E(\varepsilon) = \mathbf{0}$
- $cov(\mathbf{f}, \varepsilon) = \mathbf{0}$
- $var(\mathbf{f}) = \Delta$ definit positif
- $var(\varepsilon) = \psi = \begin{bmatrix} \psi_1 & 0 & \cdots & 0 \\ 0 & \psi_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \psi_p \end{bmatrix}$, dengan $\psi_i > 0$

Dengan asumsi di atas maka dari persamaan (2) dapat diturunkan bahwa

$$var(\mathbf{x}) = \Sigma = \mathbf{L}\Delta\mathbf{L}' + \psi$$

Karena **L** dan **f** keduanya tidak diketahui persamaan model (2) setara dengan menuliskan

$$\begin{aligned} \mathbf{x} &= \mu + \mathbf{L}\Delta^{1/2}\Delta^{-1/2}\mathbf{f} + \varepsilon \\ &= \mu + \mathbf{L}^*\mathbf{f}^* + \varepsilon \end{aligned}$$

dimana $\mathbf{L}^* = \mathbf{L}\Delta^{1/2}$ dan $\mathbf{f}^* = \Delta^{-1/2}\mathbf{f}$. Dengan bentuk model ini maka matriks ragam-peragam bagi vektor **x** adalah

$$var(\mathbf{x}) = \Sigma = \mathbf{L}^*\mathbf{L}^{*\prime} + \psi$$

Tanpa mengurangi makna untuk kasus secara umum, kita asumsikan bahwa *var*(**f**)=**I**$_k$, berupa matriks identitas, sehingga menghasilkan

$$var(\mathbf{x}) = \Sigma = \mathbf{LL}' + \psi \qquad (3)$$

Model pada persamaan (2) dan asumsi persamaan (3) merupakan model faktor baku yang banyak digunakan. Tujuan analisis ini adalah menentukan **L** dan ψ sehingga asumsi pada persamaan (3) terpenuhi. Penentuan ini menjadi masalah utama dalam analisis faktor.

Perhatikan bahwa $cov(\mathbf{x}, \mathbf{f}) = \mathbf{L}$, atau $cov(x_i, f_j) = l_{ij}$. Hal ini berimplikasi bahwa peragam dari vektor acak **x** dan vektor dari faktor bersama **f** secara lengkap ditentukan oleh matriks loading faktor **L**. Juga dapat dilihat bahwa $corr(x_i, f_j) = l_{ij}/\sqrt{\sigma_{ii}} = l_{ij}$, jika $var(x_i) = \sigma_{ii} = 1$, yaitu ketika Σ berupa matriks korelasi. Pada kasus ini, matriks loading faktor tidak lain adalah matriks koefisien korelasi antara peubah asal dengan faktor bersama.

Misalkan vektor berukuran px1, $\mathbf{l}_i$ dan $\mathbf{l}_j$ adalah baris ke-i dan ke-j dari matriks **L**. Maka untuk $i \neq j$,

$$\sigma_{ij} = cov(x_i, x_j) = \mathbf{l}_i'\mathbf{l}_j = l_{i1}l_{j1} + l_{i2}l_{j2} + \ldots + l_{ik}l_{jk}$$

dan

$$\begin{aligned}\sigma_{ii} &= var(x_i) = \mathbf{l}_i'\mathbf{l}_i + \psi_i \\ &= l^2_{i1} + l^2_{i2} + \ldots + l^2_{ik} + \psi_i \\ &= h^2_i + \psi_i\end{aligned}$$

dengan $h^2_i = \mathbf{l}_i'\mathbf{l}_i$. Jadi ragam dari x_i diuraikan menjadi dua komponen ragam, yaitu = h^2_i dan ψ_i, yang masing-masing berpadanan dengan faktor bersama dan faktor khusus. Besaran ψ_i adalah kontribusi faktor

khusus ε_i yang disebut ragam khusus, sedangkan h^2_i adalah kontribusi faktor bersama dan disebut komunalitas ragam bersama. Lebih lanjut, l^2_1 adalah kontribusi faktor bersama pertama terhadap ragam bersama, l^2_2 adalah kontribusi faktor bersama kedua terhadap ragam bersama, dan seterusnya.

Ilustrasi.

Misalkan diketahui matirks korelasi berikut

$$\rho_{4x4} = \begin{bmatrix} 1.00 & 0.51 & 0.35 & 0.20 \\ 0.51 & 1.00 & 0.21 & 0.06 \\ 0.35 & 0.21 & 1.00 & 0.35 \\ 0.20 & 0.06 & 0.35 & 1.00 \end{bmatrix}$$

Dapat diperksa bahwa ρ = **LL' +** ψ, dengan memilih matriks loading faktor

$$\mathbf{L} = \begin{bmatrix} 0.8 & -0.1 \\ 0.6 & -0.3 \\ 0.5 & 0.5 \\ 0.3 & 0.4 \end{bmatrix}$$

dan matriks ragam khusus

$$\psi = \begin{bmatrix} 0.35 & 0 & 0 & 0 \\ 0 & 0.55 & 0 & 0 \\ 0 & 0 & 0.50 & 0 \\ 0 & 0 & 0 & 0.75 \end{bmatrix}$$

Sehingga ρ memiliki struktur faktor yang dihasilkan dengan k = 2 faktor. Komunalitas dai x_1 adalah $h^2_1 = l^2_{11} + l^2_{12} = (0.8)^2 + (-0.1)^2 = 0.65$. Ragam khusus untuk x_1 adalah $\psi_1 = 0.35$. Telah diperlihatkan bahwa $h^2_1 + \psi_1 = 1 = \sigma_{11}$. Hal yang serupa juga dapat dilakukan untuk x_2, x_3, dan x_4. Selanjutnya $corr(x_1, f_1) = 0.8$, $corr(x_1, f_2) = -0.1$, $corr(x_2, f1) = 0.6$, dan seterusnya.

Umumnya, penguraian secara tepat matriks korelasi ρ ataupun matriks ragam peragam Σ menajadi **LL' +** ψ tidak selalu mungkin diperoleh. Persamaan ρ = **LL' +** ψ dalam faktor analisis hanya terpenuhi pada kondisi tertentu berkaitan dengan nilai p dan k. Selalu, kita menginginkan banyaknya parameter yang tak diketahui pada model faktor lebih sedikit daripada yang di dalam matriks Σ. Perhatikan bahwa Σ pada bagian kiri persamaan Σ = **LL' +** ψ memiliki $p(p+1)/2$ unsur yang berbeda, sedangkan disebelah kiri **L** memiliki pk unsur dan ψ memiliki p unsur, sehingga totalnya $pk+p$ parameter. Jadi untuk model faktor diperlukan bahwa

$$\frac{p(p+1)}{2} - (pk + p) > 0 \text{ atau } p > 2k.$$

Jika $p > 2k$, solusi tepat untuk persamaan Σ = **LL' +** ψ, dalam bentuk **L** dan ψ memungkinkan pada kasus sederhana. Sebagai contoh $p = 3$ dan $k = 1$

$$x_1 = l_1 f + \varepsilon_1$$

$$x_2 = l_2 f + \varepsilon_2$$

$$x_3 = l_3 f + \varepsilon_3$$

maka

$$\mathbf{\Sigma} = \mathbf{LL'} + \boldsymbol{\psi} = \begin{bmatrix} l_1 \\ l_2 \\ l_3 \end{bmatrix} \begin{bmatrix} l_1 & l_2 & l_3 \end{bmatrix} + \begin{bmatrix} \psi_1 & 0 & 0 \\ 0 & \psi_2 & 0 \\ 0 & 0 & \psi_3 \end{bmatrix}$$

Sehingga

$$\begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{bmatrix} = \begin{bmatrix} l_1^2 & l_1 l_2 & l_1 l_3 \\ l_2 l_1 & l_2^2 & l_2 l_3 \\ l_3 l_1 & l_3 l_2 & l_3^2 \end{bmatrix} + \begin{bmatrix} \psi_1 & 0 & 0 \\ 0 & \psi_2 & 0 \\ 0 & 0 & \psi_3 \end{bmatrix}$$

menyelesaikan persamaan di atas untuk mendapatkan l_i dan ψ_i, i = 1, 2, 3, menghasilkan

$$l_1^2 = \frac{\sigma_{13}\sigma_{12}}{\sigma_{23}}, \qquad l_2^2 = \frac{\sigma_{23}\sigma_{12}}{\sigma_{13}}, \qquad l_3^2 = \frac{\sigma_{13}\sigma_{23}}{\sigma_{12}}$$

(4)

dan

$$\psi_1 = \sigma_{11} - l_1^2, \qquad \psi_2 = \sigma_{22} - l_2^2, \qquad \psi_3 = \sigma_{33} - l_3^2$$

(5)

Jadi paling tidak pada kasus khusus ini, solusi tepatnya dapat dengan mudah diperoleh. Tetapi pada beberapa kasus, solusi tepat ini tidak wajar. Misalnya saja jika matriks korelasi yang terlibat adalah sebagai berikut

$$\rho = \begin{bmatrix} 1.00 & 0.75 & 0.95 \\ 0.75 & 1.00 & 0.85 \\ 0.95 & 0.85 & 1.00 \end{bmatrix}$$

Menggunakan formula-formula yang diberikan pada persamaan (4) dan (5) di atas akan didapatkan l^2_1 = 0.838, l^2_2 = 0.671, l^2_3 = 1.077, ψ_1 = 0.162, ψ_2 = 0.329 dan ψ_3 = -0.077. Padahal ψ_3 adalah ragam dari peubah acak (yaitu ε_3), sehingga tidak boleh negatif. Dengan kata lain, solusi yang didapatkan tidak wajar. Situasi seperti ini dalam banyak literatur disebut sebagai solusi tak wajar atau *kasus Heywood*. Kita akan membahas kasus ini pada sub-bab lain.

Model faktor pada persamaan (2) tidak unik karena dua pasang (**L, f**) dan (**L*, f***) yang berbeda menghasilkan struktur matriks ragam peragam Σ yang sama. Misalkan saja Γ adalah sembarang matriks ortogonal berukuran *k x k*. Maka model pada persamaan (2) dapat ditulis ulang sebagai

$$\begin{aligned} \mathbf{x} &= \mu + \mathbf{L}\Gamma\Gamma'\mathbf{f} + \varepsilon \\ &= \mu + \mathbf{L}^* \mathbf{f}^* + \varepsilon \end{aligned}$$

dengan $\mathbf{L}^* = \mathbf{L}\Gamma$ dan $\mathbf{f}^* = \Gamma'\mathbf{f}$. Perhatikan bahwa $E(\mathbf{f}^*) = E(\Gamma'\mathbf{f}) = \Gamma' E(\mathbf{f}) = \mathbf{0}$, dan $var(\mathbf{f}^*) = var(\Gamma'\mathbf{f}) = \Gamma'\ var(\mathbf{f})\ \Gamma = \Gamma'\ \Gamma = \mathbf{I}_k$. Sehingga sembarang transformasi ortogonal terhadap **f** akan menghasilkan struktur peragam yang sama untuk Σ, yaitu

$$\begin{aligned} \Sigma &= \mathbf{LL}' + \psi \\ &= \mathbf{L}\,\Gamma\,\Gamma'\,\mathbf{L}' + \psi \\ &= \mathbf{L}^*\,\mathbf{L}^{*\prime} + \psi \end{aligned}$$

Lebih jauh, komunalitas h^2_i masih sama karena $\mathbf{LL}' = \mathbf{L}^*\mathbf{L}^{*\prime}$. Walaupun komunalitas dan struktur peragam tidak berubah, tapi besarnya loading faktor sangat tergantung pada matriks transformasi ortogonal Γ.

Syarat Keunikan

Permasalahan ketidakunikan dari loading faktor sebenarnya dapat memberikan keuntungan tersendiri, karena semua kemungkinan loading faktor saling berkaitan melalui transformasi ortogonal sehingga kita dapat memilih matriks loading faktor **L** tertentu yang selain memenuhi persamaan (3) juga memiliki sifat-sifat yang lebih bermakna. Dalam banyak hal, seringkali kita menginginkan **L** adalah matriks yang memenuhi sifat bahwa **L'ψL** merupakan matriks diagonal dengan unsur positif, atau

$$\mathbf{L'\psi L} = \begin{bmatrix} \delta_1 & 0 & \cdots & 0 \\ 0 & \delta_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \delta_k \end{bmatrix}, \text{ dengan } \delta_i > 0,\ i = 1, 2, \ldots, k$$

Relevansi dari tambahan syarat ini dapat dijelaskan sebagai berikut. Misalkan **f** adalah peubah normal ganda-*k* dengan vektor rataan **0** dan matriks ragam peragam $\mathbf{I}_k$. Misalkan pula sebaran bersyarat **x** terhadap **f** adalah peubah normal ganda-*p* dengan vektor rataan $\boldsymbol{\mu} + \mathbf{Lf}$ dan matriks ragam-peragam ψ. Maka dapat ditunjukkan bahwa sebaran bersyarat **f** terhadap **x** adalah normal ganda-k dengan vektor rataan $\mathbf{L'\Sigma^{-1}(x - \mu)}$ dan matriks ragam peragam $[\mathbf{L'\psi^{-1}L} + \mathbf{I_k}]^{-1}$. Syarat yang disebutkan pada persamaan (6) memastikan bahwa matriks loading faktor **L** yang akhirnya diperoleh berdasarkan **x** menghasilkan komponen **f** yang saling bebas.

Kendala yang diberikan pada persamaan (6) mengurangi $k(k-1)/2$ batasan linear (unsur non-diagonal matriks simetrik $\mathbf{L'\psi^{-1}L}$ semuanya 0) pada jumlah parameter model faktor. Dengan demikian, syarat baru agar analsis faktor memiliki makna adalah

$$\left\{\frac{p(p+1)}{2}-\left[pk+p-\frac{k(k-1)}{2}\right]\right\} \geq 0 \qquad (7)$$

Sebagai contoh, dari pertidaksamaan (7) untuk $k = 3$, agar bermakna maka dibutuhkan $p \geq 6$, sedangkan jika $k = 4$ memerlukan $p \geq 8$.

Aspek yang paling penting pada analisis faktor adalah menduga loading faktor dan ragam khusus. Misalkan $\mathbf{x}_1, \ldots, \mathbf{x}_n$ adalah contoh acak dari sebaran ganda-p dengan $E(\mathbf{x}) = \boldsymbol{\mu}$ dan $var(\mathbf{x}) = \boldsymbol{\Sigma}$. Analisis ini terdiri atas dua bagian:

- mengidentifikasi struktur
- menduga parameter

Jika model faktor yang sesuai telah diidentifikasi dan parameter-parameter yang terlibat telah diduga, berikutnya adalah memberikan makna yang tepan dan menginterpretasi loading faktor hasil dugaan. Loading faktor yang diduga umumnya dirotaasi (ditransformasi menggunakan transformasi ortogonal), karena ketidakunikan solusi, untuk membantu menginterpretasikannya.

Identifikasi struktur yang lebih sederhana adalah menentukan k ($< p$) yang tepat sehingga model gaktor pada persamaan (2) terpenuhi. Hampir selalu, penentuan banyaknya k ini dilakukan secara subjektif, meskipun pada beberapa kasus, misalkan padapenggunaan meode pendugaan kemungkinan maksimum, pendekatan sistematik untuk menguji hipotesis dimungkinkan untuk dilakukan. Ide-ide pemilihan banyaknya komponen utama pada analisis komponen utama juga dapat digunakan untuk menentukan banyaknya faktor bersama, yaitu k.

Menyelesaikan masalah analisis faktor secara tepat adalah pekerjaan yang sulit. Inilah alasan banyaknya pendekatan yang berbeda-beda

dan sudut pandang yang dikemukakan ahli-ahli dalam menghadapi masalah ini. Untuk memudahkan diskusi, kami memabgai berbagai metode menjadi dua kategori besar (a) metode non-iteratif, dan (b) metode iteratif.

Seperti pada kasus analisis komponen utama, titik awal kerja pada analisis faktor adalah matriks ragam peragam contoh atau matriks korrelasi contoh dari data yang didapatkan. Semua metode yang akan dibahas berikut ini, dapat dilakukan baik untuk matriks ragam-peragam maupun untuk matriks korelasi. Karena sebagian besar metode tidak tergantung pada skala maka pada analisis faktor umumnya digunakan matriks korelasi. Walaupun begitu, ada juga metode yang tergantung pada skala, artinya penggunaan matriks korelasi dan matriks ragam-peragam akan menghasilkan output yang berbeda.

8.3. Metode Pendugaan Non-Iteratif

Pada pembahasan-pembahasan berikut, kita akan selalu menggunakan matriks korelasi sebagai dasar analisis. Metode yang termasuk dalam kelompok non-iteratif ini antara lain: metode komponen utama (*principal component method*), metode faktor utama (*principal factor method*), analsis citra (*image analysis*), dan analisis faktor kanonik non-iteratif Harris (*Harris' non-iterative canonical factor analysis*).

8.3.1. Metode Komponen Utama

Metode komponen utama pada analisis faktor adalah metode yang paling sederhana. Misalkan **R** adalah matriks korelasi contoh berukuran *p x p*. Karena matriks **R** adalah simetrik dan definit positif maka dapat dituliskan sebagai

$$\mathbf{R} = \Gamma\Lambda\Gamma'$$

Dengan Λ adalah diag($\lambda_1, \ldots, \lambda_p$), dan $\lambda_1 \geq \ldots \geq \lambda_p > 0$ adalah akar ciri matriks **R**, serta $\Gamma\Gamma' = \Gamma'\Gamma = \mathbf{I}_p$, dengan Γ adalah matriks ortogonal *p x p* yang kolom-kolomnya adalah vektor ciri matriks **R** , yaitu $\Gamma_1, \ldots, \Gamma_p$ yang berpadanan dengan vektor ciri $\lambda_1, \ldots, \lambda_p$. Misalkan *k* adalah banyaknya komponen utama yang dipilih menggunakan kriteria tertentu, misalnya banyaknya komponen utama minimum yanga mampu menerangkan persentase keragaman total. Kita mendefinisikan matriks $\hat{\mathbf{L}}$ berukuran *p x k* sebagai

$$\hat{\mathbf{L}} = \left[\sqrt{\lambda_1}\mathbf{\Gamma}_1 \,|\cdots|\, \sqrt{\lambda_k}\mathbf{\Gamma}_k\right] \qquad (10)$$

maka, **R** didekati dengan

$$\hat{\mathbf{L}}\hat{\mathbf{L}}' = \sum_{i=1}^{k} \lambda_i \mathbf{\Gamma}_i \mathbf{\Gamma}_i'$$

dimana Γ_i adalah kolom ke-*i* pada matriks Γ.

Jadi $\hat{\mathbf{L}} = (\hat{l}_{ij})$ yang diberikan pada persamaan (10) merupakan penduga matriks loading faktor **L**. Matriks diagonal ragam khusus ψ diduga dengan $\hat{\boldsymbol{\psi}}$, yaitu matriks diagonal yang unsurnya diambil dari **R** - $\hat{\mathbf{L}}\hat{\mathbf{L}}'$

$$\hat{\boldsymbol{\psi}} = \begin{bmatrix} 1-h_1^2 & 0 & \cdots & 0 \\ 0 & 1-h_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1-h_p^2 \end{bmatrix}$$

dimana $h^2_i = \sum_{i=1}^{k} l_{ij}^2$, i = 1, 2, …, *p*

Dengan demikian kita memiliki model *k*-faktor dengan **L** diduga oleh $\hat{\mathbf{L}}$ dan ψ diduga dengan $\hat{\psi}$, sehingga diperoleh pendekatan bagi **R** adalah

$$\mathbf{R} \approx \hat{\mathbf{L}}\hat{\mathbf{L}}' + \hat{\psi}$$

Dalam prakteknya kita harus melakukan validasi model, sehingga sebelum menerima $\hat{\mathbf{L}}$ dan $\hat{\psi}$ sebagai penduga akhir, perlu dihitung matriks sisaan

$$\mathbf{Res} = \mathbf{R} - (\hat{\mathbf{L}}\hat{\mathbf{L}}' + \hat{\psi})$$

dan besaran dari unsur-unsurnya diperhatikan menggunakan ukuran statistik tertentu atau secara intuitif.

Matriks sisaan **Res** selalu memiliki unsur diagonal nol. Pada kasus yang ideal, **Res** = **0**. Dengan demikian, dalam bahasa intuitif, jika unsur non-diagonal juga dekat dengan nilai 0 maka penduga $\hat{\mathbf{L}}$ dan $\hat{\psi}$, dianggap cukup bagus dan dapat diterima. Hal ini memicu munculnya statistik untuk menilai mutu dari pendugaan.

Secara formal, jika *k* terpilih sehingga $\sum_{i=1}^{k} \lambda_i$ yang merupakan keragaman total yang diterangkan oleh *k* buah vektor ciri besar (mendekati *p*) maka $\sum_{i=k+1}^{p} \lambda_i$ akan kecil (mendekati 0). Bisa ditunjukkan bahwa jumlah kuadrat dari unsur-unsur matriks sisaan Res selaku kurang

dari atau sama dengan $\sum_{i=k+1}^{p} \lambda_i$ (Rao, 1964). Sehingga jika $\sum_{i=k+1}^{p} \lambda_i$ kecil, ini juga menunjukkan bahwa matriks sisaan itu kecil . Atau besaran

$$\frac{\sum_{i=k+1}^{p} \lambda_i}{\sum_{i=1}^{p} \lambda_i} = \frac{\sum_{i=k+1}^{p} \lambda_i}{p}$$

dapat dijadikan ukuran kebaikan suai dari model faktor. Semakin kecil nilainya mengindikasikan kebaikan suai yang tinggi. Perlu ditegaskan bahwa jika yang digunakan bukan matriks korelasi melainkan matriks ragam peragam, maka $\sum_{i=1}^{p} \lambda_i$ tidak selalu sama dengan *p*.

Ilustrasi.

Kembali menggunakan data yang digunakan oleh Spearman pada sub-bab sebelumnya, kita bekerja dengan menggunakan matriks korelasi

$$\mathbf{R} = \begin{matrix} C \\ F \\ E \\ M \\ D \\ Mu \end{matrix} \begin{bmatrix} 1.00 & 0.83 & 0.78 & 0.70 & 0.66 & 0.63 \\ 0.83 & 1.00 & 0.67 & 0.67 & 0.65 & 0.57 \\ 0.78 & 0.67 & 1.00 & 0.64 & 0.54 & 0.51 \\ 0.70 & 0.67 & 0.64 & 1.00 & 0.45 & 0.51 \\ 0.66 & 0.65 & 0.54 & 0.45 & 1.00 & 0.40 \\ 0.63 & 0.57 & 0.51 & 0.51 & 0.40 & 1.00 \end{bmatrix}$$

Penguraian spektral matriks di atas mendapatkan nilai akar ciri sebagai berikut: 4.1028, 0.6191, 0.5117, 0.3570, 0.2704, dan 0.1389. Jika digunakan kriteria bahwa faktor yang diambil berpadanan dengan akar ciri yang lebih dari 1, maka kita dapatkan k = 1, sehingga bentuk matriks loading faktornya berupa vektor saja, dan diperoleh $\boldsymbol{l}$ = (l_1, …, l_p), dengan p = 6 adalah

$$\boldsymbol{l} = (0.9365, 0.8940, 0.8420, 0.8024, 0.7426, 0.7207)$$

Karena k = 1, maka penduga komunalitasny $h^2_i = l^2_i$, i = 1, …, 6 berpadanan dengan enam peubah yang terlibat sehingga diperoleh 0.8771, 0.7992, 0.7090, 0.6467, 0.5514, dan 0.5194. Sebagai contoh untuk $h^2_1 = l^2_1 = (0.9365)^2 = 0.8771$. Penduga untuk total komunalitas adalah

$$\sum_{i=1}^{p} h_i^2 = \sum_{i=1}^{p} l_i^2 = 4.1028$$

Nilai ini hanya 68.4% dari total keragaman 6 peubah asal. Ragam khusus untuk masing-masing enam peubah diperoleh dari $\psi_i = 1 - h^2_i$, i = 1, …, 6, yaitu

0.1229, 0.2008, 0.2910, 0.3533, 0.4486, 0.4806

Informasi-informasi di atas dapat dirangkum pada tabel berikut

	C	F	E	M	D	Mu
Loading Faktor	0.9365	0.8940	0.8420	0.8042	0.7426	0.7207
Komunalitas	0.8771	0.7992	0.7090	0.6467	0.5514	0.5194
Ragam khusus	0.1229	0.2008	0.2910	0.3533	0.4486	0.4806
RMS	0.0353	0.0548	0.0706	0.0813	0.0986	0.0895

Matriks sisaan **Res** = **R** – ($\hat{\mathbf{L}}\hat{\mathbf{L}}' + \hat{\boldsymbol{\psi}}$) memberikan indikasi seberapa bagus model faktor mampu mengepas data. Berdasar data di atas didapatkan

Res = (*res*$_{ij}$) =

$$\begin{bmatrix} 0.0000 & -0.0072 & -0.0086 & -0.0531 & -0.0354 & -0.0450 \\ -0.0072 & 0.0000 & -0.0828 & -0.0489 & -0.0138 & -0.0743 \\ -0.0086 & -0.0828 & 0.0000 & -0.0371 & -0.0853 & -0.0969 \\ -0.0531 & -0.0489 & -0.0371 & 0.0000 & -0.1472 & -0.0696 \\ -0.0354 & -0.0138 & -0.0853 & -0.1472 & 0.0000 & -0.1352 \\ -0.0450 & -0.0743 & -0.0969 & -0.0696 & -0.1352 & 0.0000 \end{bmatrix}$$

Salah satu ukuran untuk menilai kebaikan suai adalah menggunakan RMS_overall yaitu akar kuadrat tengah dari seluruh unsur non diagonal matriks **Res**, atau

$$\text{RMS_overall} = \sqrt{\frac{1}{p(p-1)}\sum_{i=1}^{p}\sum_{j\neq i}^{p} res_{ij}^{2}}$$

Nilai RMS_overall untuk kasus di atas adalah 0.0748. Meskipun tidak adal keseragaman aturan unruk menenrukan batasan dari RMS_overall ini, namun beberapa pendapat menyatakan bahwa nilai 0.05 dapat digunakan. Artinya jika diperoleh RMS_overall yang kurang dari 0.05 dengan banyaknya faktor bersama paling sedikit, itulah yang diambil sebgai model terbaik berdasarkan kriteria ini.

Untuk melakukan evaluasi terhadap masing-masing peubah, dapat juga didapatkan RMS di setiap peubah, yang didapatkan menggunakan formula

$$\text{RMS}_i = \sqrt{\frac{1}{p-1}\sum_{j \neq i} res_{ij}^2}$$

Sehingga misalkan untuk peubah C didapatkan

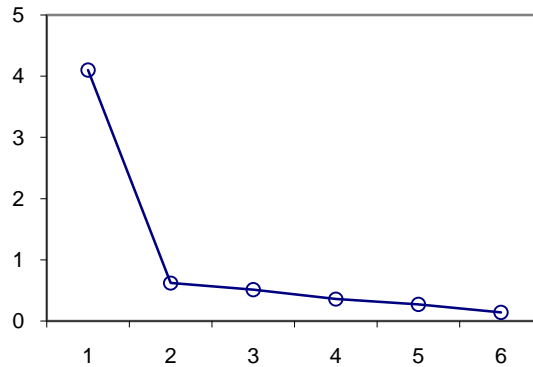
$$\text{RMS}_C = \text{RMS}_1 = \sqrt{\frac{1}{p-1}\sum_{j \neq i} res_{1j}^2}$$

$$= \sqrt{\frac{1}{5}\left[(-0.0072)^2 + (-0.0086)^2 + (-0.0531)^2 + (-0.0354)^2 + (-0.0450)^2\right]}$$

$$= 0.0353$$

Mungkin ada faktor bersama lain yang kalau dimasukkan ke dalam model akan meningkatkan persentase keragaman total yang diterangkan melalui komunalitas dari faktor-faktor. Namun, memasukkan faktor tambahan ini tidak selalu dapat menurunkan RMS.

Petunjuk berapa banyak faktor yang diikutsertakan dalam model faktor kadang-kadang dilakukan dengan menggunakan grafik yang disebut *SCREE DIAGRAM*. Grafik ini menampilkan nilai akar ciri dari matriks korelasi lawan *k*, *k* = 1, ..., *p*. Dengan grafik ini, *k*, banyaknya faktor, dipilih sedemikian rupa sehingga gradien dari grafik tersebut curam di sebelah kirinya dan sangat landai ke sebelah kanan. Diagram ini serupa dengan yang digunakan pada analisis komponen utama.

Scree diagram pada kasus di atas adalah sebagai berikut

Dari diagram di atas diperoleh bahwa minimal dua faktor diperlukan pada model. Jika dipilih k = 2 akan diperoleh RMS_overall 0.0771, yang mengidikasikan bahwa model faktor dengan dua faktor bersama tidak memperbaiki hasil. Bahkan nilai RMS_overall ini lebih tinggi daripada ketika k = 1. Selanjutnya, total keragaman yang mampu diterangkan oleh dua buah faktor ini adalah

$$\frac{\lambda_1 + \lambda_2}{\lambda_1 + \cdots + \lambda_6} = 78.7\%$$

Jika dicoba menggunakan k = 3 dan k = 4 berturut-turut akan diperoleh RMS_overall sebesar 0.0541 dan 0.0384, yang memperlihatkan perbaikan berarti dibandingkan ketika k = 1 atau k = 2. Komunalitas yang dihasilkan dengan menggunakan model tersebut adalah 87% dan 93%. Sebagai bahan perbandingan, tabel berikut menyajikan RMS setiap peubah dan RMS_overall untuk berbagai k.

	C	F	E	M	D	Mu	Overall
k = 1	0.0353	0.0548	0.0706	0.0813	0.0986	0.0895	0.0748
k = 2	0.0346	0.0510	0.0706	0.0836	0.0867	0.1114	0.0771
k = 3	0.0343	0.0510	0.0715	0.0747	0.0511	0.0213	0.0541
k = 4	0.0322	0.0513	0.0279	0.0425	0.0466	0.0212	0.0384

8.3.2. Metode Faktor Utama

Metode ini didasarkan pada penggunaan metode komponen utama terhadap matriks korelasi semu (atau matriks ragam-peragam semu), yaitu diperoleh dengan mengganti unsur diagonal dengan komunalitas.

Misalkan struktur faktor yang diberikan pada persamaan (3) $\mathbf{R} = \mathbf{LL'} + \psi$, berlaku untuk suatu nilai *k*, seperti sebelumnya *k* adalah banyaknya faktor. Maka matriks korelasi semu $\mathbf{R} - \psi = \mathbf{LL'}$ merupakan matriks simetrik semidefinit positif dengan pangkat *k*. Tujuan kita adalah menduga **L**. Menggunakan penguraian spektral terhadap $\mathbf{R} - \psi$ menghasilkan

$$\mathbf{R} - \psi = \Gamma\Lambda\Gamma'$$

dimana $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_k, 0, \ldots, 0)$, dan $\lambda_1 \geq \ldots \geq \lambda_k > 0$ adalah akar ciri dari matriks korelasi semu, dan Γ adalah matriks vektor ciri padanannya. Dengan demikian matriks **L** dapat diperoleh melalui

$$\hat{\mathbf{L}} = \left[\sqrt{\lambda_1}\mathbf{\Gamma}_1 | \cdots | \sqrt{\lambda_k}\mathbf{\Gamma}_k\right]$$

dengan Γ_i adalah kolom ke-*i* dari matriks Γ. Dalam prakteknya untuk matriks **R** yang diketahui dan $\psi = \hat{\psi}$, $\mathbf{R} - \hat{\psi}$ mungkin saja berpangkat penuh. Jadi dengan menggunakan $\hat{\Gamma}$ sebgai penduga Γ, dan $\hat{\Lambda}$ sebagai penduga Λ, melalui penguraian

$$\mathbf{R}_a = \mathbf{R} - \hat{\psi} = \hat{\Gamma}\hat{\Lambda}\hat{\Gamma}'$$

Kita menduga **L** sebagai

$$\hat{\mathbf{L}} = \left[\sqrt{\hat{\lambda}_1}\hat{\mathbf{\Gamma}}_1 | \cdots | \sqrt{\hat{\lambda}_k}\hat{\mathbf{\Gamma}}_k\right]$$

dimana $\hat{\mathbf{\Gamma}}_i$ adalah kolom ke-*i* dari matriks $\hat{\mathbf{\Gamma}}$. Sehingga kita menduga **L** dengan mendekati menggunakan $\hat{\mathbf{L}}$ yang memenuhi struktur faktor

$$\mathbf{R} \approx \hat{\mathbf{L}}\,\hat{\mathbf{L}}' + \hat{\boldsymbol{\psi}}, \text{ untuk } \psi = \hat{\boldsymbol{\psi}} \text{ yang telah ditentukan.}$$

Misalkan matriks korelasi **R** dan matriks dugaan ragam khusus $\hat{\boldsymbol{\psi}}$ telah ditentukan. Maka menggunakan $\hat{h}_i^2 = 1 - \hat{\psi}_i$, $i = 1, \ldots, p$ kita dapat memperoleh penguraian spektral

$$\mathbf{R}_a = \mathbf{R} - \hat{\boldsymbol{\psi}} = \begin{bmatrix} \hat{h}_1^2 & r_{12} & \cdots & r_{1p} \\ r_{12} & \hat{h}_2^2 & \cdots & r_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{p1} & r_{p2} & \cdots & \hat{h}_p^2 \end{bmatrix} = \hat{\mathbf{\Gamma}}\,\hat{\mathbf{\Lambda}}\,\hat{\mathbf{\Gamma}}'$$

Andaikan k' adalah banyaknya unsur diagonal $\hat{\mathbf{\Lambda}}$ yang tidak sama dengan 0, yaitu $\lambda_1, \ldots, \lambda_k$, dan $\hat{\mathbf{\Gamma}}_i, \ldots, \hat{\mathbf{\Gamma}}_{k'}$ adalah vektor ciri **R -** $\hat{\boldsymbol{\psi}}$ yang berpadanan dengannya. Meskipun **R -** ψ matriks yang definit positif dengan pangkat *k*, namun $\mathbf{R}_a = \mathbf{R} - \hat{\boldsymbol{\psi}}$ dapat memiliki akar ciri yang negatif. Kita mengambil *k'* sebgai penduga *k*. Pada metode faktor utama, $\hat{k}$ sebagai penduga *k* dipilih sehingga memenuhi $\hat{k}$ < *k'* < *k,* dan $\sum_{i=1}^{\hat{k}} \hat{\lambda}_i$ dekat dengan penduga totaal komunalitas $\sum_{i=1}^{p} h_i^2$ = *tr*(**R -** $\hat{\boldsymbol{\psi}}$). Dengan dasar ini matriks **L** diduga dengan

$$\hat{\mathbf{L}} = \left[\sqrt{\hat{\lambda}_1}\hat{\mathbf{\Gamma}}_1 \mid \cdots \mid \sqrt{\hat{\lambda}_k}\hat{\mathbf{\Gamma}}_{\hat{k}} \right]$$

Pendugaan **L** yang dinyatakan di atas ditentukan tergantung pada $\hat{\boldsymbol{\psi}}$ yang ditentukan atau pada komunalitas $\hat{h}_i^2$. Artinya pada saat menggunakan metode ini, kita harus menentukan matriks $\hat{\boldsymbol{\psi}}$ atau menentukan komunalitasnya. Ada banyak cara menentukan hal ini, dan kita akan bahas beberapa di antaranya.

8.3.2.1. Pendugaan menggunakan Squared Multiple Correlation (SMC)

Misalkan $\mathbf{R}^{-1} = (r^{ij})$ adalah matriks kebalikan dari matriks korelasi **R**. Pendekatan yang dapat digunakan untuk menduga ψ_i adalah $\hat{\psi}_i = 1/r^{ii}$, $i = 1, \ldots, p$. Hal ini setara dengan melakukan pendugaan terhadap komunalitas h^2_i dengan $\hat{h}_i^2 = 1 - \hat{\psi}_i = 1 - 1/r^{ii} = SMC_i$, yaitu sebesar kuadrat dari koefisien korelasi ganda antara x_i, komponen ke-i pda vektor peubah **x**, dengan $(p - 1)$ komponen **x** lainnya. Matriks korelasi semu $\mathbf{R}_a$ pada persamaan (12), dengan mengganti h^2_i dengan $\hat{h}_i^2$ sekarang digunakan untuk melakukan analisis faktor. Total komunalitas tidak lain diduga menggunakan $\sum_{i=1}^{p} SMC_i$.

8.3.2.2. Pendugaan menggunakan nilai mutlak korelasi yang terbesar

Metode intuitif lain yang dapat digunakan untuk menduga komunalitas h^2_i adalah menggunakan

$$r_{max}^{(i)} = \max_{i \neq j} | r_{ij} |$$

yaitu nilai terbesar dari nilai mutlak koefisien korelasi pada unsur non-diagonal baris ke-i dari matriks **R**. Sehingga

$$\hat{h}_i^2 = r_{max}^{(i)} \;, i = 1, \ldots, p$$

8.3.2.3. Pendugaan menggunakan Adjusted Squared Multiple Correlation (ASMC).

Cureton dan D'Agostino (1983) menunjukkan bahwa $\sum_{i=1}^{p} r_{max}^{(i)}$ adalah penduga total komunalitas $\sum_{i=1}^{p} h_i^2$ yang lebih baik daripada $\sum_{i=1}^{p} SMC_i$. Sehingga mereka menyarankan dilakukan penyesuaian (pengkoreksian) terhadap kuadrat dari korelasi ganda untuk menduga h^2_i sehingga $\sum_{i=1}^{p} h_i^2$ masih didugan menggunakan $\sum_{i=1}^{p} r_{max}^{(i)}$. Yang mereka usulkan adalah sebagai berikut

$$\hat{h}_i^2 = \text{ASMC}_i = \frac{\sum_{j=1}^{p} r_{max}^{(j)}}{\sum_{j=1}^{p} \text{SMC}_j} \text{SMC}_i \quad , i = 1, \ldots, p$$

Dari persamaan di atas, kondisi bahwa $\sum_{i=1}^{p} h_i^2$ diduga dengan $\sum_{i=1}^{p} r_{max}^{(i)}$ tetap terpenuhi.

8.3.3. Analisis Citra (*Image Analysis*)

Meskipun pada analisis faktor secara jelas sisebutkan di dalam modelnya bagaimana faktor bersama dan faktor khusus berhubungan, identifikasi terhadap bagian-bagian ini seringkali dilakukan menggunakan trial and error. Untuk mengatasi ketidakmampuan dari analisis faktor menyatakan secara eksplisit faktor bersama dan faktor khusus pada sebuah peubah, Guttman (1953) memperkenalkan sebuah metode yang sekarang populer dengan nama analisis citra.

Hal utama yang mencolok pada metode ini adalah kemampuan secara eksplisit mendefinikan bagian faktor bersama pada peubah. Guttman mengatakan 'bagian faktor bersama pada sebuah peubah adalah bagian yang dapat diprediksi menggunakan korelasi linear ganda dari semua peubah lain pada gugus peubah tersebut.' Bagian bersama dari peubah ini disebut sebagai *image* dari peubah tersebut berdasarkan peubah lain. Bagian khusus dari peubah tersebut adalah sisaan yang tidak dapar diprediksi. Bagian khusus ini disebut sebagai *anti-image* dari peubah. Dengan menggunakan terminologi analisis regresi, image dan anti-image dapat dijelaskan sebagai berikut.

Misalkan **x** adalah vektor acak berukuran *p x 1* yang telah dibakukan sehingga matriks ragam-peragamnya tidak lain adalah matriks korelasi **R**, dan diasumsikan non-singular. Jika model regresi

$$x_i = \beta_1 x_1 + \ldots + \beta_{i-1} x_{i-1} + \beta_i x_{i+1} + \ldots + \beta_{p-1} x_p + \varepsilon_i$$

digunakan untuk meprediksi x_i, maka penduga kuadrat terkecil bagi x_i adalah $\hat{x}_i$ yaitu

$$\hat{x}_i = \mathbf{R}_{12(i)} \mathbf{R}_{22(i)}^{-1} \mathbf{x}_{2(i)}$$

dimana $\mathbf{x} = (x_i, \mathbf{x}'_{2(i)})$, $\mathbf{x}_{2(i)}$ adalah vektor yang memuat seluruh peubah kecuali x_i, serta matriks **R** dipartisi menjadi

$$\mathbf{R} = \begin{bmatrix} 1 & \mathbf{R}_{12(i)} \\ \mathbf{R}_{21(i)} & \mathbf{R}_{22(i)} \end{bmatrix}.$$

Misalkan sisaan $\hat{\varepsilon}_i$ diperoleh dari $\hat{\varepsilon}_i = x_i - \hat{x}_i$. Maka x_i dapat dituliskan dalam bentuk bagian bersama dan bagian khusus, atau *image* dan *anti-image* menjadi

$$x_i = \hat{x}_i + \hat{\varepsilon}_i \;,\; i = 1, \ldots, p$$

Dalam bentuk vektor, *p* buah persamaan di atas dituliskan sebagai

$$\mathbf{x} = \hat{\mathbf{x}} + \hat{\boldsymbol{\varepsilon}} \qquad (13)$$

dimana $\mathbf{x} = (x_1, \ldots, x_p)'$, $\hat{\mathbf{x}} = (\hat{x}_1, \ldots, \hat{x}_p)'$ dan $\hat{\boldsymbol{\varepsilon}} = (\hat{\varepsilon}_1, \ldots, \hat{\varepsilon}_p)'$. Selanjutnya telah dibuktikan bahwa

$$\hat{\mathbf{x}} = (\mathbf{I} - \mathbf{M}\mathbf{R}^{-1})\mathbf{x}$$

dengan **M** adalah matriks diagonal yang unsur diagonalnya diambil dari unsur diagonal matriks $\mathbf{R}^{-1}$. Melalui pendefinisian $\mathbf{W} = \mathbf{I} - \mathbf{M}\mathbf{R}^{-1}$, kita dapat menuliskan ulang persamaan (13) menjadi

$$\mathbf{x} = \mathbf{W}\mathbf{x} + (\mathbf{I} - \mathbf{W})\mathbf{x}$$

$$= \textit{image} + \textit{anti-image}$$

Perhatikan bahwa matriks ragam peragam dari *image* adalah

$$var(\mathbf{W}\mathbf{x}) = \mathbf{W}\mathbf{R}\mathbf{W}' \qquad (14)$$

$$= (\mathbf{I} - \mathbf{M}\mathbf{R}^{-1})\,\mathbf{R}\,(\mathbf{I} - \mathbf{M}\mathbf{R}^{-1})'$$

$$= \mathbf{R} + \mathbf{M}\mathbf{R}^{-1}\mathbf{M} - 2\mathbf{M}$$

Perlu menjadi catatan bahwa unsur diagonal utama pada matriks ragam peragam adalah kuadrat dari koefisien korelasi berganda dari sebuah peubah dengan peubah-peubah lainnya. Pada analisis faktor metode image dan menggunakan matriks korelasi **R**, matriks ragam-peragam dari image pada persamaan (14) dihitung terlebih dahulu. Kemudian berikutnya matriks inilah yang dijadikan bahan ekstraksi menggunakan metode komponen utama untuk menghasilkan faktor bersama. Faktor-faktor yang berpadanan dengan akar ciri dari **WRW'** yang lebih dari atau sama dengan satu umumnya yang digunakan.

Ilustrasi.

Jika digunakan data Spearman seperti pada ilustrasi sebelumnya, kita akan dapatkan matriks **W** dan **WRW'** sebagai berikut

W =

$$\begin{bmatrix} 0.0000 & 0.5465 & 0.6106 & 0.2898 & 0.4040 & 0.4622 \\ 0.3708 & 0.0000 & 0.0026 & 0.2835 & 0.3568 & 0.1363 \\ 0.3043 & 0.0019 & 0.0000 & 0.2249 & 0.0681 & 0.0172 \\ 0.1179 & 0.1700 & 0.1835 & 0.0000 & -0.0922 & 0.1071 \\ 0.1432 & 0.1864 & 0.0484 & -0.0803 & 0.0000 & -0.0512 \\ 0.1460 & 0.0635 & 0.0109 & 0.0832 & -0.0456 & 0.0000 \end{bmatrix}$$

dan

WRW' =

$$\begin{bmatrix} 0.8142 & 0.7284 & 0.6665 & 0.6461 & 0.5841 & 0.5441 \\ 0.7284 & 0.7261 & 0.6693 & 0.5923 & 0.5523 & 0.5327 \\ 0.6665 & 0.6693 & 0.6271 & 0.5561 & 0.5146 & 0.5036 \\ 0.6461 & 0.5923 & 0.5561 & 0.5431 & 0.4921 & 0.4611 \\ 0.5841 & 0.5523 & 0.5146 & 0.4921 & 0.4756 & 0.4268 \\ 0.5441 & 0.5327 & 0.5036 & 0.4611 & 0.4268 & 0.4118 \end{bmatrix}$$

Akar ciri dari matriks **WRW'** ini adalah 3.4698, 0.0698, 0.0399, 0.0166, 0.0025, dan 0.0002. Jika faktor yang digunakan hanya yang berpadanan dengan akar ciri yang lebih dari 1 maka hanya satu faktor bersama yang terpilih. Matriks (tapi berupa vektor, karena hanya satu faktor) loading faktor yang merupakan vektor ciri yang berpadanan dengan akar ciri pertama adalah

(0.9724, 0.9073, 0.8329, 0.7778, 0.7057, 0.6720)'

Nilai RMS_overall-nya adalah 0.0507, dan komunalitas dari setiap peubah berturut-turut adalah 0.9456, 0.8233, 0.6937, 0.6050, 0.4980, dan 0.4516.

8.3.4. Analisis Faktor Kanonik Non-Iteratif Harris

Rao (1955) memperkenalkan analisis faktor kanonik untuk menjawab pertanyaan 'peubah faktor mana yang berhubungan paling erat dengan peubah teramati?' Solusi untuk permasalahan ini ada kaitannya dengan analisis korelasi kanonik terhadap peubah faktor hipotetik dengan peubah teramati.

Mengacu kepada model yang diberikan pada persamaan (2), matriks peragam antara **x** dan **f** adalah matriks berukuran *(p+k)* x *(p+k)*, yaitu

$$var\left(\begin{bmatrix} \mathbf{x} \\ \mathbf{f} \end{bmatrix}\right) = \begin{bmatrix} \mathbf{\Sigma} & \mathbf{L} \\ \mathbf{L'} & \mathbf{I} \end{bmatrix}$$

Dengan mengasumsikan bahwa peubah-peubah telah dibakukan maka bentuk dari Σ adakan menjadi matriks korelasi ρ. Sehingga notasi ρ akan digunakan untuk menggantikan Σ. Tujuannya adalah menduga **L** dan ψ, dimana ρ = **LL' +** ψ. Kuadrat dari koefisien korelasi kanonik antara **x** dan **f** diperoleh dari akar persamaan berikut

$$|\mathbf{LL'} - v\rho| = 0$$

Misalkan matriks ρ diduga menggunakan matriks korelasi contoh **R** dan misalkan juga bahwa dugaan awal bagi ψ , $\hat{\psi}$ telah tersedia. Maka kita menginginkan

$$\mathbf{LL'} \approx \mathbf{R} - \hat{\psi}$$

Dengan mengguanakan $\hat{\mathbf{L}}$ sebagai penduga **L**, dan $\hat{\mathbf{L}}\,\hat{\mathbf{L}}' = \mathbf{R} - \hat{\boldsymbol{\psi}}$, maka persamaan seberlumnya dapat ditulis ulang menjadi

$$| \mathbf{R} - \hat{\boldsymbol{\psi}} - v\mathbf{R} | = 0$$

atau

$$| \hat{\boldsymbol{\psi}}^{-1/2} \mathbf{R} \hat{\boldsymbol{\psi}}^{-1/2} - \lambda \mathbf{I} | = 0$$

dengan $\lambda = 1/(1-v)$

Dalam penerapannya bukan **R** yang digunakan tetapi matriks korelasi semu $\mathbf{R}_a$, yaitu dengan mengganti unsur diagonal **R** dengan $\hat{h}_i^2$. Sehingga persamaan yang berlaku sekarang adalah

$$| \hat{\boldsymbol{\psi}}^{-1/2} \mathbf{R}_a \hat{\boldsymbol{\psi}}^{-1/2} - \lambda \mathbf{I} | = 0$$

Karena $\mathbf{R}_a$ tidak selalu berupa matriks definit positif, maka dengan penggantian ini akar ciri $\hat{\boldsymbol{\psi}}^{-1/2} \mathbf{R}_a \hat{\boldsymbol{\psi}}^{-1/2}$, yaitu λ dapat bernilai negatif. Misalkan saja $\hat{\lambda}_1 \geq \ldots \geq \hat{\lambda}_k > 0$ adalah akar ciri $\hat{\boldsymbol{\psi}}^{-1/2} \mathbf{R}_a \hat{\boldsymbol{\psi}}^{-1/2}$ yang bernilai positif dan $\boldsymbol{\Gamma}_1$ adalah matriks berukuran *p x k* yang merupakan matriks vektor ciri padanannya. Kemudian disusun matriks $\hat{\boldsymbol{\Lambda}}_1^{1/2}$ sebagai

$$\hat{\boldsymbol{\Lambda}}_1^{1/2} = \begin{bmatrix} \sqrt{\hat{\lambda}_1} & 0 & \cdots & 0 \\ 0 & \sqrt{\hat{\lambda}_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sqrt{\hat{\lambda}_k} \end{bmatrix}$$

serta mendefinisikan $\hat{\mathbf{L}} = \hat{\psi}^{-1/2}\Gamma_1 \hat{\boldsymbol{\Lambda}}_1^{1/2}$ sebagai penduga bagi **L**. Jika mau, kita dapat memeperoleh penduga yang baru bagi ψ sebagai $\hat{\psi} = \mathbf{R} - \hat{\mathbf{L}}\,\hat{\mathbf{L}}'$ dan melajutkan secara iteratif prosedur ini hingga tercapai kestabilan penduga komunalitas. Proses ekstraksi faktor ini besifat scale-invariant. Lebih jelasnya misalkan dilakukan transformasi penskalaan **y** = **Dx**, dimana **D** adalah matriks diagonal yang semua tandanya sama. Solusi bagi **L** tidak berubah meskipun data **x** diganti dengan data baru y. Hal ini karena solusi yang dihasilkan dari ekstraksi matriks

$$(\mathbf{D}\hat{\psi}\mathbf{D}')^{-1/2}\,\mathbf{D}\mathbf{R}_a\mathbf{D}'(\mathbf{D}\hat{\psi}\mathbf{D}')^{-1/2}$$

yang berpadanan dengan **y** sama dengan yang diperoleh menggunakan $\hat{\psi}^{-1/2}\,\mathbf{R}_a\,\hat{\psi}^{-1/2}$ pada **x**. Konsekuensinya penggunaan $\hat{\boldsymbol{\Sigma}}$ maupun **R** menghasilkan solusi yang identik.

Harris (1962) telah menunjukkan hubungan yang sangat penting antara analisis faktor kanonik yang dikembangkan Rao dengan analisis image yang dikembangkan Guttman. Menggunakan hasil dari Guttman bahwa jika *k* cukup kecil dibandingkan *p* maka $[\text{diag}(\mathbf{R}^{-1})]^{-1} \doteq \psi$ jika $p \to \infty$, dia mengusulkan suatu algoritma non-iteratif bagi analisis faktor kanonik dari Rao dengan cara mengekstraksi akar ciri dan vektor ciri dari matriks

$$[\text{diag}(\mathbf{R}^{-1})]^{1/2}\ \mathbf{R}_a\ [\text{diag}(\mathbf{R}^{-1})]^{1/2}$$

8.4. Metode Pendugaan Iteratif

Banyak metode ekstraksi iteratif yang termotivasi oleh penggunaan metode kemungkinan maksimum. Sebagai alternatifnya, ada juga prosedur yang telah disusun tanpa mengasumsikan sebaran tertentu. Pada bagian ini kita akan membahas emapt metode iteratif yaitu: metode kemungkinan maksimum (*maximum likelihood method*), metode kuadrat terkecil tak-terboboti (*unweighted least squares*, ULS), metode komponen utama iteratif (*iterative principal component method*), dan analisis faktor alpha (*alpha factor analysis*).

8.4.1. Metode Kemungkinan Maksimum

Metode kemungkinan maksimum (MKM) pada analisis faktor dilakukan awalnya dengan mengasumsikan sebaran normal ganda diperkenalkan oleh Lawley (1940) dan algoritmanya kemudian dikembangkan oleh Joreskog (1967, 1977). MKM mengasumsikan bahwa matriks ragam peragam atau matriks korelasi semua peubah bersifat non-singular. Karena metode ini *scale-invariant* maka kita dapat memperoleh solusi menggunakan salah satu matriks tersebut. Misalkan $\mathbf{x}_1, \ldots, \mathbf{x}_n$ adalah contoh acak dari populasi yang menyebar normal ganda-*p* dengan vektor rataan μ dan matriks ragam-peragam Σ, yang diasumsikan memiliki struktur peragam $\Sigma = \mathbf{LL'} + \psi$. Permasalahan analisis faktor dapat dipandang sebagai pencarian penduga kemungkinan maksimum bagi $\mathbf{L}$ dan ψ.

Sudah diketahui bahwa rataan vektor contoh $\bar{\mathbf{x}}$ merupakan penduga kemungkinan maksimum bagi μ. Untuk mendapatkan penduga kemungkinan maksimum bagi $\mathbf{L}$ dan ψ, kita gunakan matriks ragam-peragam contoh $\mathbf{S}$ sebagai penduga awal bagi Σ. Karena contoh diasumsikan bersal dari populasi yang menyebar normal ganda, maka

$(n\text{-}1)\mathbf{S}$ memiliki sebaran Wishart dengan derajat bebas $(n\text{-}1)$ dan nilai harapan **S** adalah Σ. Fungsi kepekatan peluang bagi **S** adalah L(**S**) diberikan oleh

$$L(S) = c.|\mathbf{\Sigma}|^{-\frac{n-1}{2}} |\mathbf{S}|^{\frac{n-1}{2}-\frac{p+1}{2}} e^{\frac{n-1}{2} tr(\mathbf{\Sigma}^{-1}\mathbf{S})}$$

dengan *c* adalah konstanta. Sehingga log-likelihood dari **L** dan ψ , jika Σ = **LL'** + ψ adalah

$$ln\ c - \left(\frac{n-1}{2}\right) tr\{[(\mathbf{LL'}+\boldsymbol{\psi})^{-1}\mathbf{S}] - ln|\mathbf{LL'}+\boldsymbol{\psi})^{-1}\mathbf{S}| \qquad (16)$$

Penduga kemungkinan maksimum bagi **L** dan ψ diperoleh dengan memaksimumkan fungsi (16) dengan kendala $k(k\text{-}1)/2$ persyaratan kenunikan (Johnson & Wichern, 1998).

8.4.1.1. Penentuan banyaknya faktor bersama

Terdapat beberapa metode yang dapat digunakan untuk menentukan banyaknya faktor bersama, antara lain:

1. Uji Nisbah Kemungkinan (*likelihood ratio test*)

Salah satu keuntungan penggunaan metode kemunkinan maksimum pada analisis faktor adalah adanya cara untuk menguji hipotesis bahwa faktor sebanyak *k* adalah pilihan yang tepat, dengan *k* adalah bilangan bulat tak diketahui. Formalisasi ini menghilangkan perdebatan tentang kesubjektifan penentuan banyaknya faktor bersama. Hipotesis nol yang diuji pada uji nisbah kemungkinan ini adalah

$H_0 : \Sigma = \mathbf{LL'} + \psi$, $r(\mathbf{L}) = k$ diketahui

Misalkan $\hat{\mathbf{L}}$, $\hat{\boldsymbol{\psi}}$, dan $\hat{\boldsymbol{\Sigma}} = \hat{\mathbf{L}}\hat{\mathbf{L}}' + \hat{\boldsymbol{\psi}}$ adalah penduga kemungkinan maksimum bagi **L**, ψ dan Σ, jika H_0 benar, maka nilai maksimum untuk log dari fungsi kemungkinannya adalah

$$\ln L_{H0} = c^* - \left(\frac{n-1}{2}\right)\{tr[(\hat{\boldsymbol{\Sigma}})^{-1}\mathbf{S}] - \ln|\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{S}|\}$$

Jika Σ merupakan matriks definit positif tak berstruktur, maka penduga kemungkinan maksimum bagi Σ adalah **S**, sehingga nilai maksimum untuk log dari fungsi kemungkinan menjadi

$$\ln L_{H0} = c^* - \left(\frac{n-1}{2}\right)\{tr[\mathbf{S}^{-1}\mathbf{S}] - \ln|\mathbf{S}^{-1}\mathbf{S}|\}$$

$$= c^* - \frac{n-1}{2}p$$

Jadi jika H_0 benar dan n sangat besar, maka statistik uji nisbah kemungkinan adalah

$$-2\ln\lambda = -2\ln\left(\frac{\mathcal{L}_{H_0}}{\mathcal{L}}\right)$$

$$= (n-1)\{tr(\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{S}) - \ln|\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{S}| - p\}$$

$$= (n-1)\{tr[(\hat{\mathbf{L}}\hat{\mathbf{L}}'+\boldsymbol{\psi})^{-1}\mathbf{S}] - \ln|(\hat{\mathbf{L}}\hat{\mathbf{L}}'+\boldsymbol{\psi})^{-1}\mathbf{S}| - p\}$$

$$= (n-1)F(\hat{\mathbf{L}}, \hat{\boldsymbol{\psi}})$$

dimana $F(\mathbf{L}, \psi) = \{tr[(\mathbf{L}\mathbf{L}'+\boldsymbol{\psi})^{-1}\mathbf{S}] - \ln|(\mathbf{L}\mathbf{L}'+\boldsymbol{\psi})^{-1}\mathbf{S}| - p\}$ mendekati sebaran khi-kuadarat dengan derajat bebar ½ $[(p-k)^2 - (p+k)]$.

Nilai derajat bebas di atas dihitung berdasarkan selisih antara banyaknya parameter yang diduga jika $\mathbf{\Sigma}$ tidak berstruktur dan banyaknya paremeter jika H_0 benar. Ada sebanyak $p(p+1)/2$ parameter jika tidak ada kendala pada $\mathbf{\Sigma}$ serta ada $(pk + p)$ parameter jika hipotesis nol benar. Namun adanya kendala sebanyak $k(k - 1)/2$ pada persyaratan keunikan mengurangi jumlah parameter yang diduga jika H_0 benar. Sehingga derajat bebas dari uji ini adalah

$$\frac{p(p+1)}{2} - \left\{pk + p - \frac{k(k-1)}{2}\right\} = ½\ [(p-k)^2 - (p+k)].$$

Pendugaan L dan ψ. Untuk mendapatkan penduga kemungkinan maksimum bagi **L** dan ψ, Joreskog (1977) mengusulkan proses iterasi peminimuman dua tahap. Pertama, untuk matriks ψ tertentu, minimum dari

$$F(\mathbf{L}, \psi) = tr[(\mathbf{LL'} + \boldsymbol{\psi})^{-1}\mathbf{S}] - \ln|(\mathbf{LL'} + \boldsymbol{\psi})^{-1}\mathbf{S}| - p$$

terhadap **L** didapatkan, misalkan saja nilai minimumnya adalah $F(\hat{\mathbf{L}}, \psi)$. Selanjutnya $F(\hat{\mathbf{L}}, \psi)$ diminimumkan terhadap ψ. Algoritma ini terus dilakukan secara iterasi sehingga diperoleh dugaan bagi **L** dan ψ yang membuat minimum global dari $F(\mathbf{L}, \psi)$ tercapai.

Rao (1955) menunjukkan bahwa solusi faktor yang dihasilkan oleh metode kemungkinan maksimum dengan menagasumsikan sebaran normal ganda sama dengan hasil pendugaan menggunakan analisis faktor kanonik yang dikembangkannya. Dengan demikian, untuk menduga loading faktor dan ragam khusus menggunakan metode kemungkinan maksimum, tidak

diperlukan asumsi kenormalan ganda. Namun untuk pengujian hipotesis, asumsi kenormalan ini tetap diperlukan.

2. Kriteria Informasi Akaike (Akaike's information criterion)

Sebagai alternatif untuk menentukan banyak faktor yang sesuai dapat digunakan besaran *Akaike's information criterion* (AIC), selain melakukan pengujian. AIC pertama kali diperkenalkan oleh Akaike (1973) untuk menduga banyaknya parameter dalam sebuah model, dan penerapannya pada analisis faktor dibahas pada tahun 1987. Dengan melibatkan k buah faktor dalam model, maka matriks ragam peragam Σ dapat dituliskan $\Sigma = \mathbf{L}_k\mathbf{L}'_k + \psi$, dengan $\mathbf{L}_k$ adalah matriks loading faktor berukuran *p x k*. Sehingga log dari fungsi kemungkinan yang berpadanan dengan model k-faktor ini berdasarkan pada data contoh acak dari populasi normal ganda-p, $\mathbf{x}_1, \ldots, \mathbf{x}_n$ adalah

$$ln\ L(k) = c + \frac{n}{2}\left[\ln|\mathbf{L}_k\mathbf{L}_k' + \mathbf{\Psi}| + tr((\mathbf{L}_k\mathbf{L}_k' + \mathbf{\Psi})^{-1}\mathbf{S}_n)\right],$$

dimana $\mathbf{S}_n = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})'$. Maka statistik AIC untuk model dengan *k* faktor didefinisikan sebagai

$$AIC(k) = -2\ ln\ L(k) + [2p(k+1) - k(k-1)]$$

Model berfaktor *k* dengan *k* adalah nilai yang berpadanan dengan AIC(*k*) yang paling kecil dianggap sebagai model yang paling baik. Selain AIC, beberapa orang juga menggunakan kriteria lain yang disebut *Schwarz's Bayesian Criterion* (SBC) yang kegunaannya sama.

8.4.2. Metode Kuadrat Terkecil Tak-Terboboti (Unweighted Least Squares Method, ULS)

Joreskog (1977) mengembangkan algoritma untuk menduga **L** dan ψ secara iteratif, yaitu meminimumkan fungsi kuadrat terkecil

$$F_U(\mathbf{L}, \psi) = ½\, tr(\mathbf{S} - \Sigma)^2 = ½\, tr[(\mathbf{S} - \Sigma)\ (\mathbf{S} - \Sigma)]$$

terhadap **L** dan ψ, dimana $\Sigma = \mathbf{LL}' + \psi$, dan **LL'** adalah matriks diagonal berukuran *k x k*. Solusi permasalan ini diperoleh menggunakan algoritma yang disebut ULS. Meskipun metode ini tidak *scale-invariant*, tapi metode ini memiliki kelebihan karena tidak mengasumsikan matriks ragam peragam (atau matriks korelasi) harus non-singular. Secara intinya, algoritma untuk metode ULS ini sama dengan algoritma iteratif dua tahap seperti yang dibahas pada bagian sebelumnya.

Solusi menggunakan ULS pada analisis faktor ini juga ekuivalen dengan solusi yang diperoleh jika menggunakan metode lain yang kurang terkenal yaitu metode jumlah kuadrat sisaan minimum (*minimum residual sum of squares*, MINRES) yang dikembangkan oleh Harman (1977). Karena itulah metode MINRES ini tidak didiskusikan di sini.

8.4.3. Metode Faktor Utama Beriterasi

Berharap mendapatkan penduga yang lebih baik untuk matriks loading faktor dan ragam khusus, kadang-kadang kita perlu melakukan iterasi terhdap metode faktor utama yang dibahas sebelumnya sampai perubahan pada penduga komunalitas akhir sudah lebih kecil daripada nilai yang ditentukan sebelumnya. Itulah ide dasar dari metode faktor utama beriterasi ini.

Misalkan $\hat{\mathbf{L}}^{(i)}$ adalah penduga matriks loading faktor pada iterasi ke-*i*. Maka penduga bagi ψ pada iterasi berikutnya adalah

$$\hat{\psi}^{(i+1)} = diag(\mathbf{S} - \hat{\mathbf{L}}^{(i)}\, \hat{\mathbf{L}}^{(i)}\,')$$

dan selanjutnya melakukan ekstraksi komponen utama terhadap matriks $\hat{\boldsymbol{\psi}}^{(i+1)}$ untuk mendapatkan $\hat{\mathbf{L}}^{(i+1)}$. Proses iterasi terus dilanjutkan hingga terpenuhi kriteria kekonvergenan tertentu. Matriks korelasi contoh **R** juga dapat digunakan untuk menggantikan **S**, namun perlu dicatat bahwa hasilnya belum tentu sama.

Prosedur iteratif yang digambarkan di atas setara dengan meminimumkan

$$tr[(\mathbf{S} - \Sigma)^2] = \sum_i \sum_j (s_{ij} - \sigma_{ij})^2$$

dengan kendala

$$\Sigma = \mathbf{LL}' + \psi$$

dalam dua tahap. Dengan demikian, metode faktor utama beriterasi ini setara dengan metode ULS kecuali pada metode ULS ada tambahan kendala yaitu matriks **LL**' merupakan matriks diagonal.

8.5. Kasus Heywood

Dalam faktor analisis, tujuan utama yang dilakukan dapat digambarkan sebagai proses meminimumkan penyimpangan matriks ragam peragam (atau matriks korelasi) contoh **S** dengan matriks ragam-peragam $\Sigma = \mathbf{LL}' + \psi$, untuk mendapatkan $\hat{\mathbf{L}}$ dan $\hat{\boldsymbol{\psi}}$ dengan kendala ψ bersifat definit positif. Soulsi yang diperoleh akdang-kadang berupa solusi yang tak wajar, seperti misalkan diperoleh dugaan ragam khusus bernilai negatif. Kasus seperti ini dikenal dengan *kasus Heywood*. Secara definisi, kasus Heywood terjadi jika selama proses iterasi paling tidak diperoleh dugaan ragam khusus $\hat{\psi}_i$ bernilai negatif.

Karena $\hat{\psi}_i$ merupakan penduga ragam, maka dugaan seperti itu tidak layak, dan dengan demikian tidak dapat diterima.

Kasus Heywood dapat terjadai pada praktek karena matriks $\Sigma = LL' + \psi$ mungkin bersifat definit positif meskipun ada satu atau lebih unsur ψ yang bernilai o atau negatif. Hal ini dapat dilihat pada konsturksi matriks definit positif yang dikemukakan oleh Heywood (1931) berikut ini.

Andaikan kita memilih $l_i, l_2, \ldots, l_p$ yang memenuhi konsisi $l_1 > 0$, $0 < l_i < 1$, $i = 2, \ldots, p$ dan

$$l_1^2 \leq 1 + \frac{1}{\dfrac{l_2^2}{1-l_2^2} + \cdots + \dfrac{l_p^2}{1-l_p^2}}$$

serta misalkan $\psi_1, \psi_2, \ldots, \psi_p$ diperoelh sehingga $l^2_1 + \psi_1 = 1$, $l^2_2 + \psi_2 = 1$, …, $l^2_p + \psi_p = 1$. Maka matriks yang bentuknya

$$\begin{bmatrix} l_1^2 + \psi_1 & l_1 l_2 & \cdots & l_1 l_p \\ l_2 l_1 & l_2^2 + \psi_2 & \cdots & l_2 l_p \\ \vdots & \vdots & \ddots & \vdots \\ l_p l_1 & l_p l_2 & \cdots & l_p^2 + \psi_p \end{bmatrix}$$

bersifat definit positif. Perhatikan bahwa ψ_1 dapat bernilai ≤ 0. Konstruksi matriks definit positif di atas berbeda dengan solusi analisis faktor karena ada salah satu ψ, yaitu ψ_1 yang bernilai < 0, sehingga didapatkan $l^2_1 \geq 1$, sedangkan pada faktor analisis harus bernilai kurang dari 1. Sehingga kasus Heywood dapat menghasilkan solusi yang tak wajar. Namun perlu diperhatikan bahwa solusiyang tak wajar tidak selau berupa kasus Heywood. Menggunakan studi empiris, beberapa

penyebab yang dididentifikasi menimbulkan solusi tak wajar antara lain:

- ukuran contoh terlalu kecil
- banyaknya variabel terlalu sedikit
- terlalu banyak faktor bersama yang diekstraksi
- faktor bersama yang diekstraksi terlalu sedikit
- pendugaan awal komunalitas kurang tepat
- model faktor kurang tepat
- adanya pencilan

Terjadinya solusi yang tak wajar tidak harus dipandang dari sisi negatifnya. Jika memungkinkan, penyebab dari kejadian ini harus diidentifikasi untuk selanjutnya dilakukan perbaikan. Hal-hal yang mungkin dapat memperbaiki solusi anatara lain adalah

- abaikan masalah ini jika penduganya tidak berbeda nyata dibandingkan solusi yang wajar
- buang peubah yang bermasalah
- perbaiki solusi tak wajar dengan nilai yang dapat diterima
- gunakan contoh yang lebih besar
- perbanyak peubah yang terlibat
- gunakan kendala pertidaksaman utnuk mencegak solusi tak wajar
- identifikasi dan buang data pencilan sebelum menerapakan analisis faktor.

8.6. Rotasi Faktor

Telah disinggung sebelumnya bahwa matriks loading faktor **L** yang diperoelh tergantung pada pemilihan matriks transformasi ortogonal Γ, dalam pengertian bahwa jika **L** adalah matriks loading faktor maka **L*** = **L** Γ juga merupakan matriks loadingfaktor asalkan Γ ortogonal. Ketidakunikan ini dapat menimbulkan kebingungan pada pengguna analisis faktor. Jika matriks loading faktor yang diperoleh sulit diinterpretasikan maka disarankan untuk mentransformasi matriks

tersebut dengan mengalikan matriks ortogonal terhadapnya sehingga interpretasi yang bermakna menggunakan matriks yang baru itu memungkinkan. Proses ini dikenal dengan sebutan rotasi daktor, karena dengan melakukan hal ini bukan faktor asal **f** yang digunakan melainkan faktor yang dirotasi $\mathbf{f}^* = \Gamma\,\mathbf{f}$.

Secara umum, rotasi faktor dilakukan sedemikian rupa sehingga faktor yang sudah dirotasi memiliki sedikit saja peubah dengan nilai mutlak loading yang besar, sedangkan sisanya kecil atau nol. Pola seperti ini akan memudahkan pengguna memberikan interpretasi terhadap faktor yang terbentuk. Sebagai teladan misalkan pada hasil analisis faktor didapatkan bahwa hanya tiga peubah pertama saja yang memiliki loading besar pada faktor bersama pertama. Maka faktor bersama tersebut dapat diinterpretasikan sebagai kombinasi linear hanya dari tiga peubah asal tadi.

Dalam transformasi, beberapa literatur juga menyebutkan transformasi linear tak-ortogonal, atau dikenal juga sebagai *rotasi oblique*.

8.6.1. Rotasi Ortogonal

Beberapa metode untuk menentukan matriks ortogonal yang sesuai dilakukan dengan merotasi faktor telah diusulkan. Metode-metode itu mengoptimalkan fungsi tujuan tertentu untuk memperoleh matriks ortogonal.

8.6.1.1. Rotasi Quartimax

Salah satu pendekatan yang terkenal adalah merancang matriks transformasi ortogonal sehinga ragam yang dihitung dari kuadrat loading faktor hasil transformasi mencapai maksimum. Yaitu, jika **L** adalah matriks loading faktor yang ingin ditranformasi mengunakan matriks ortogonal Γ menjadi $\mathbf{L}^* = \mathbf{L}\,\Gamma$, maka Γ dipilih sehingga

$$\frac{1}{pk}\sum_i\sum_j l_{ij}^{*4} - \left(\frac{1}{pk}\sum_i\sum_j l_{ij}^{*2}\right) = \frac{1}{pk}\sum_i\sum_j l_{ij}^{*4} - \left(\frac{1}{pk}\sum_i h_i^{*2}\right)$$

(19)

mencapai maksimum. Karena komunalitas $h_i^{*2} = \sum_{j=1}^{k} l_{ij}^{*2} = \sum_{j=1}^{k} l_{ij}^{2} = h_i^2$, i = 1, ..., p seluruhnya adalah konstanta dan tidak tergantung pada transformasi Γ, maka memaksimumkan (19) sama saja dengan memaksimumkan unsur pertamanya saja yaitu $\sum_i \sum_j l_{ij}^{*4}$. Transformasi ortogonal denga tujugan memperoleh Γ yang memaksimumkan $\sum_i \sum_j l_{ij}^{*4}$ dikenal sebagai rotasi quartimax.

8.6.1.2. Rotasi varimax kasar

Kaiser (1958) mengusulkan bahwa ragam dari kuadrat loading yang berpadanan pada setiap kolom dihitung dan dijumlahkan, yaitu sebesar

$$\sum_{j=1}^{k}\left\{\frac{1}{p}\sum_{i=1}^{p}(l_{ij}^{*2})^2 - \left[\frac{1}{p}\sum_{i=1}^{p} l_{ij}^{*2}\right]^2\right\} \qquad (20)$$

dimaksimumkan untuk mendapatkan matriks transformasi ortogonal Γ. Pendekatan menggunakan hal ini disebut *rotasi varimax kasar*.

8.6.1.3. Rotasi varimax

Prosedur *varimax* yang merupakan rotasi paling sering digunakan pada aplikasi merupakan transformasi ortogonal yang diperoleh dengan cara memaksimumkan

$$\sum_{j=1}^{k}\left\{\frac{1}{p}\sum_{i=1}^{p}(\frac{l_{ij}^{*2}}{h_i})^2-\left[\frac{1}{p}\sum_{i=1}^{p}\frac{l_{ij}^{*2}}{h_i}\right]^2\right\}, \qquad (21)$$

yang sebetulnya adalah versi pembakuan dari fungsi tujuan pada (20).

8.6.1.4. Rotasi-rotasi lainnya

Harman (1976) tertarik untuk memaksimumkan kmbinasi linear dari fungsi tujuan pada rotasi quartimax dan varimax kasar. Dia mengusulkan untuk memaksimumkan

$$\frac{1}{p}\sum_{j=1}^{k}\left\{\sum_{i=1}^{p}l_{ij}^{*4}-\frac{\gamma}{p}\left(\sum_{i=1}^{p}l_{ij}^{*2}\right)^2\right\} \qquad (22)$$

dengan γ adalah sebuah konstanta. Kriteria yang diberikan oleh (22) sangat umum, karena pemilihan γ yang berbeda akan berimplikasi didapatkan transformasi ortogonal yang berbeda juga. Memilih $\gamma = 0$ sama saja dengan melakukan trnasformasi quartimax, $\gamma = 1$ akan sama dengan rotasi varimax kasar, $\gamma = k/2$ akan menghasilkan *rotasi equimax* (Saunders, 1962), dan $\gamma = p(k-1)/(p+k-2)$ akan menghasilkan *rotasi parsimax* (Crawford, 1967). Rotasi equimax dan parsimax tidak kita diskusikan disini.

8.6.2. Rotasi Oblique

Kadangkala, bahkan setelah dilakukan transformasi ortogonal terhadap matriks loading faktor, faktor yang dihasilkan masih sulit untuk diinterpretasikan. Pada kasus seperti ini, *rotasi oblique* tertentu disarankan untuk dilakukan. Idenya adalah menemukan transformasi terhadap sumbu loading faktor asal sehingga sumbu hasil transformasi melalui kelompok loading faktor lebih dekat daripada yang dihasilkan

oleh transformasi ortogonal. Jadi, untuk matriks loading faktor tertentu **L** ingin diperoleh matriks **T** berukuran *k x k*, yang tidak harus matriks ortogonal sehingga matriks **L*** = **LT** memiliki interpretasi yang lebih bermakna daripada **L**.

8.6.2.1. Rotasi HK

Harris dan Kaiser (1964) menyatakan bahwa matriks T yang dicari dapat jadi merupakan fungsi dari beberapa matriks ortogonal dan matriks diagonal sederhana. Sebagai misal, matriks transformasi T mungkin berbentuk

$$\mathbf{T} = \Gamma_2 \mathbf{D}_2 \Gamma_1 \mathbf{D}_1$$

dengan Γ_2 dan Γ_1 adalah matriks ortogonal, serta $\mathbf{D}_1$ dan $\mathbf{D}_2$ adalah matriks diagonal. Salah saku kelompok rotasi oblique Harris-Kaisser adalah dengan mengambil $\Gamma_2 = \mathbf{I}_k$, sehingga

$$\mathbf{T} = \mathbf{D}_2 \Gamma_1 \mathbf{D}_1$$

Dalam hal ini $\mathbf{D}_2$ dan Γ_1 dipilih terlebih dahulu, dan selanjutnya $\mathbf{D}_1$ dipilih untuk melakukan penskalaan yang sesuai, sehingga matriks korelasi faktornya adalah $\Phi = \mathbf{D}_1^{-1} \Gamma_1' \mathbf{D}_2^{-1} \Gamma_1 \mathbf{D}_1^{-1}$. Nantinya Φ adalah sedemikian sehingga matriks yang terpilih pada pemfaktoran dapat dinyatakan sebagai $(\mathbf{LT})\ \Phi(\mathbf{LT})'$. Harris dan Kaiser meyarankan untuk pertama kali menggunakan metode faktor utama untuk mendapatkan **L**. Misalkan Λ adalah matriks diagonal dari akar ciri matriks korelasi semu yang digunakan pada metode faktor utama, maka pemilihan matriks $\mathbf{D}_2$ dan Γ_1 yang disarankan adalah $\mathbf{D}_2 = \Lambda^{-1/2}$ dan Γ_1 adalah matriks transformasi quartimax.

Jika menggunakan transformasi ortogonal $\mathbf{L}^* = \mathbf{L}\ \Gamma$ mempertahankan sifat bahwa antaf faktor tidak berkorelasi, namun trnasformasi oblique $\mathbf{L}^* = \mathbf{LT}$ menyebabkan adanya korelasi antar faktor. Matriks korelasi

antar faktor adalah (**T'T**)$^{-1}$ Jika yang digunakan sebagai matriks yang diekstraksi adalah matriks korelasi maka korelasi antara peubah asal dengan faktor bersama, tidak lain adalah **L**. Namun jika digunakan rotasi oblique **L*** = **LT**, maka matriks korelasinya adalah **L**(**T**')$^{-1}$. Nilai korelasi ini berguna untuk studi lebih mendalam pada pemilihan faktor, namun matriks pola faktor yang terotasi yang seharusnya digunakan untuk interpretasi faktor.

8.6.2.2. Rotasi Promax

Transformasi oblique lainyya adalah *rotasi promax* yang dikemukakan oleh Hendrickson dan White (1964). Mereka mengasumsikan bahwa struktur sederhana untuk loading faktor yang dihasilkan transformasi ortogonal mirip dengan struktru sederhana yang dihasilkan rotasi oblique. Sehingga mereka memulai dengan matriks loading faktor hasil transformasi ortogonal. Transformasi ortogonal yang biasa digunakan adalah varimax. Sehingga diskusi rotasi faktor berikut ini diasumsikan menggunakan varimax.

Misalkan **L*** = **L** Γ adalah loading faktor yang ditransformasi varimax. Disusun dulu matirks **Q** = (q_{ij}) berukuran *p x k* dan disebut sebagai matriks target, yang diasumsikan memiliki struktur sederhana

$$q_{ij} = | l^{*m-1}_{ij} | \, l^{*}_{ij}$$

dengan *m* adalah bilangan asli lebih dari 1. Sehingga q_{ij} memiliki tanda yang sama dengan l^{*}_{ij} dan $|q_{ij}| = | l^{*}_{ij} |$.

Alasan mengambil nilai q_{ij} melalui pemangkatan l^{*}_{ij} dengan *m* adalah untuk mengupayakan unsur-unsurnya memiliki nilai yang lebih kecil dan mendekati nol dengan cepat. Hendrickson dan White

merekomendasikan matriks target **Q** didekati dengan transformasi oblique terhadap **L***. Rotasi promax diberikan oleh matriks

$$\mathbf{T} = \mathbf{U}\{ diag(\mathbf{U}'\mathbf{U})^{-1/2}\}$$

dengan **U** adalah matriks berukuran *k x k*, $\mathbf{U} = (\mathbf{u}_1 \mid \mathbf{u}_2 \mid \ldots \mid \mathbf{u}_k)$, dan kolom ke-*j* matriks **U** yaitu $\mathbf{u}_j$ diperoleh dengan meminimumkan bentuk kuadratik

$$(\mathbf{q}_j - \mathbf{L}^*\mathbf{u}_j)'(\mathbf{q}_j - \mathbf{L}^*\mathbf{u}_j)$$

terhadap $\mathbf{u}_j$, $j = 1, 2, \ldots, k$. Vektor $\mathbf{q}_j$ adalah kolom ke-*j* dari matriks target **Q**. Solusi dari *k* buah permasalahan minimisasi di atas dapat dinyatakan bersama sebagai

$$\mathbf{U} = (\mathbf{L}^{*\prime}\mathbf{L}^*)^{-1}\mathbf{L}^{*\prime}\mathbf{Q}.$$

8.7. Menduga Skor Faktor

Prediksi atau dugaan nilai faktor bersama yang berpadanan dengan pengamatan dengan nilai peubah asal tertentu disebut sebagai skor faktor untuk pengamatan tersebut. Nilai-nilai ini dapat digunakan sebagai data yang telah tereduksi untuk analisis statistik lanjutan. Misalkan, nilai tersebut sering digunakan untuk tujuan diagnostik. Nilai sekor dihitung setelah model faktor yang sesuai telah ditentukan. Artinya, setelah dugaan matriks loading faktor diperoleh dan rotasi yang sesuai dilakukan, baru ditentukan skor faktor untuk setiap objek (individu).

Kami menyajikan dua metodeuntuk memprediksi skor faktor. Keduanya mengasumsikan bahwa data mentah (nilai peubah asal) tersedia sehingga skor faktor dapat dihitung.

8.7.1. Metode Kuadrat Terkecil Terboboti (*Weighted Least Squares Method*)

Metode ini menentukan nilai skor faktor yang berpadanan dengan pengamatan $\mathbf{x}_i$, i = 1, …, n didapatkan menggunakan formula kuadrat kecil terboboti

$$\hat{\mathbf{f}}_i = (\hat{\mathbf{L}}'\hat{\boldsymbol{\psi}}^{-1}\hat{\mathbf{L}})^{-1}\hat{\mathbf{L}}'\hat{\boldsymbol{\psi}}^{-1}(\mathbf{x_i} - \bar{\mathbf{x}}) = \hat{\boldsymbol{\Delta}}^{-1}\hat{\mathbf{L}}'\hat{\boldsymbol{\psi}}^{-1}(\mathbf{x_i} - \bar{\mathbf{x}})$$

dimana $\bar{\mathbf{x}}$ merupakan vektor rataan contoh dari data. Penduga $\hat{\mathbf{L}}$ dan $\hat{\boldsymbol{\psi}}$ diperoleh menggunakan metode kemungkinan maksimum. Jika matriks korelasi yang digunakan (atau peubah asal telah dibakukan terlebih dahulu) untuk analisis faktor, maka

$$\mathbf{z}_i = \begin{bmatrix} 1/\sqrt{s_{11}} & 0 & \cdots & 0 \\ 0 & 1/\sqrt{s_{22}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1/\sqrt{s_{pp}} \end{bmatrix} (\mathbf{x_i} - \bar{\mathbf{x}})$$

digunanan menggantikan $(\mathbf{x_i} - \bar{\mathbf{x}})$ pada formula skor faktor di atas.

8.7.2. Metode Regresi

Metode ini menduga skor faktor untuk pengamatan yang berpadanan dengan $\mathbf{x}_i$ menggunakan formula yang sama dengan formula koefisien regresi linear, yaitu

$$\hat{\mathbf{f}}_i = \hat{\mathbf{L}}'\mathbf{S}^{-1}(\mathbf{x_i} - \bar{\mathbf{x}}), \ i = 1, \ldots, n$$

Jika matriks korelasi yang digunakan untuk ekstraksi faktor maka digunakan $\mathbf{z}_i$ seperti yang didefinisikan pada bagian sebelumnya untuk menggantikan $(\mathbf{x_i} - \bar{\mathbf{x}})$, serta **R** menggantikan posisi **S**.

8.8. Aplikasi SAS

Prosedur SAS yang digunakan untuk analisis factor adalah PROC FACTOR.

PROC FACTOR menyediakan berbagai metode untuk ekstraksi dan rotasi. Bentuk umum dari perintah PROC FACTOR adalah:

PROC FACTOR *<options>* ;

VAR *variables* ;

PRIORS *communalities* ;

RUN;

Ada tiga metode utama yang banyak digunakan dalam mendapatkan skor faktor, yaitu: Metode Kemungkinan Maksimum (maximum likelihood), Metode Kuadrat Terkecil Tak-Terboboti (unweighted least squares), dan Metode Komponen Utama. Dalam PROC FACTOR ketiganya dilakukan dengan cara memberikan perintah METHOD=opsi. Opsi yang dimaskud adalah sesuai dengan metodenya, yaitu ML, ULS dan PRIN.

Untuk pendugaan komunalitas, ada beberapa metode yang sering digunakan dan tersedia di PROC FACTOR. Dua metode yang paling sering digunakan , yaitu berdasarkan pada korelasi berganda kuadrat antar tiap peubah yang diamati dengan peubah teramati lainnya, atau berdasarkan korelasi maksimum antara tiap peubah teramati dengan peubah teramati lainnya. Metode yang umum dipakai adalah

yang berdasarkan kuadrat dari korelasi berganda (squared multiple correlation, SMC).

Berikut adalah ilustrasi penggunaan SAS untuk analisis factor.

Data yang digunakan adalah data tentang tingkat sakit yang dirasakan oleh 121 pasien yang mengeluhkan sakit di bagian muka dan rahang. Pasien-pasien tersebut disinyalir menderita gejala yang dinamakan Painful Temporomandibular Disorder atau TMD. Peubah yang diukur dari pasien tersebut didasarkan pada pertanyaan berikut:

1. apakah anda sakit saat mengunyah ? → HURTCHEW
2. apakah anda sakit saat membuka mulut ? → HURTWIDE
3. apakah rahang anda berisik ? → NOISE
4. apakah anda memiliki keluhan sakit di bagia telinga atau telinga bagian luar ? → EARPAIN
5. apakah anda punya keluahan sakit di bagian muka, rahang, mata, tenggorokan, leher ? → FACEPAIN
6. apakah sakit yang anda derita mengganggu tidur anda ? → PAINSLP
7. apakah sakit yang anda derita mengganggu kegiatan rutin anda ? → PAINROUT
8. apakah anda mengkonsumsi obat-obat penghilang rasa sakit ? → PAINTAB
9. apakah teman anda mengatakan bahwa gigi anda bergertak saat tidur ? → GRIND
10. apakah anda sadar bahwa anda sering bertopang dagu ? → CLAMPSET
11. apakah rahang anda kaku ketika bangun tidur ? → STIFFJAW
12. apakah rahang atau gigi anda sakit ketika bangun tidur ? → SOREJAW
13. apakah anda sakit kepala ketika bangun tidur ? → AMHDACHE
14. apakah rahang anda sulit digerakkan di ketika bangun tidur ? → CRACKING

Data yang didapatkan telah dituliskan dalam SAS sebagai berikut.

```
data contoh;
 input isub hurtchew hurtwide noise stiffjaw cracking
       painslp painrout paintab amhdache earpain grind
       clampset sorejaw facepain dayslost ;
 cards;
  1  2  3  3  2  3  3  3  4  5  4  3  3  3  2  4
  2  5  5  5  4  4  4  4  5  5  5  5  5  5  4  5
  3  6  7  7  6  6  4  3  4  5  5  4  4  4  4  6
  4  5  6  5  5  4  6  6  9  7  6  5  5  5  5  9
  5  3  4  5  3  3  5  4  4  6  6  4  5  4  4  5
  6  4  5  5  4  4  5  4  5  6  6  6  5  5  3  6
  7  5  6  5  4  4  3  3  4  4  6  5  5  5  4  5
  8  5  6  4  4  4  4  4  4  5  6  5  5  5  4  6
  9  3  4  5  2  4  5  6  6  7  5  5  3  4  4  5
 10  5  6  6  5  5  3  3  3  4  6  6  5  5  4  6
 11  5  7  7  5  5  4  3  4  5  5  4  4  3  4  5
 12  6  7  5  5  5  5  6  5  6  6  6  5  4  4  7
 13  5  6  5  4  5  3  3  4  5  6  5  5  5  3  5
 14  4  5  4  4  4  3  3  4  4  5  4  4  5  4  6
 15  3  5  5  3  3  4  3  4  6  5  5  4  3  3  3
 16  3  4  3  3  4  5  4  6  6  4  3  3  3  3  6
 17  4  4  4  3  3  6  6  6  7  4  3  3  3  4  5
 18  3  4  3  3  3  3  4  4  5  5  4  4  4  4  5
 19  4  6  4  3  5  4  4  4  5  6  4  4  5  5  5
 20  4  6  4  4  4  4  3  4  5  6  5  5  4  4  5
 21  5  6  4  4  5  3  3  3  5  5  5  3  4  3  5
 22  5  6  5  4  4  4  4  5  6  6  4  5  5  4  6
 23  6  7  6  5  5  2  2  3  4  6  6  5  5  3  6
 24  5  6  5  5  5  3  3  4  5  5  4  3  4  3  5
 25  5  6  5  4  6  3  5  3  4  5  5  5  4  3  5
 26  4  5  4  4  3  6  5  6  7  6  5  5  6  3  8
 27  2  4  3  1  1  4  5  4  5  6  5  4  4  2  5
 28  6  7  6  5  6  4  5  5  6  7  6  5  6  4  7
 29  5  6  6  4  4  4  3  4  5  6  5  6  4  3  5
 30  5  6  6  5  4  3  3  3  5  7  6  5  6  3  5
 31  4  5  4  4  5  4  5  5  6  6  5  5  5  2  7
 32  5  6  6  4  5  4  4  5  5  5  5  5  4  3  5
 33  4  5  5  4  4  4  5  5  6  5  3  4  3  2  5
 34  5  6  6  4  4  5  5  6  7  5  4  3  4  4  7
 35  5  6  5  4  5  4  4  5  5  5  4  6  3  3  5
 36  5  6  5  4  5  5  4  6  7  6  5  5  5  5  7
 37  4  5  4  4  5  5  5  6  7  7  6  4  5  4  6
 38  5  6  6  5  5  2  3  3  4  6  6  5  5  2  6
 39  3  4  4  3  3  4  4  4  6  5  4  3  4  3  4
 40  5  6  6  5  4  2  2  3  3  7  5  5  5  3  5
 41  4  5  5  3  4  6  6  5  7  4  3  2  3  3  6
 42  5  6  5  3  4  5  5  6  7  5  5  4  4  4  5
 43  5  6  5  4  4  4  5  5  6  5  5  5  4  4  5
 44  4  5  5  4  3  4  5  4  6  5  6  5  3  3  5
```

45	4	5	5	3	4	4	4	5	6	5	5	5	3	4	6
46	4	5	4	3	4	4	5	5	6	7	6	5	5	3	6
47	5	6	6	4	5	7	5	6	8	7	6	6	6	4	7
48	4	5	5	4	5	5	5	6	6	7	6	5	6	3	6
49	5	6	6	5	4	4	4	4	6	5	4	4	3	4	4
50	5	7	6	5	5	3	2	3	4	4	3	2	2	4	4
51	3	4	4	3	4	3	3	4	5	7	6	5	7	3	5
52	4	4	4	3	5	4	4	4	5	7	6	6	5	3	6
53	4	5	5	5	5	3	3	4	5	5	5	3	5	1	5
54	5	6	6	5	5	4	5	5	7	5	4	3	3	5	5
55	4	6	6	4	4	6	6	6	8	8	7	6	7	4	6
56	5	6	5	5	4	5	5	5	6	8	7	6	6	4	6
57	6	7	6	5	6	5	4	5	7	5	4	4	4	4	6
58	5	6	5	4	5	4	4	4	5	6	6	4	5	4	5
59	4	5	5	3	4	4	4	5	6	5	5	5	5	5	7
60	5	6	5	4	5	4	4	5	7	6	5	5	5	3	6
61	5	6	4	5	4	3	3	2	4	4	4	3	3	2	4
62	2	3	3	2	3	4	4	3	5	7	6	5	5	2	5
63	5	6	7	5	5	4	4	4	6	6	5	5	5	4	5
64	5	6	5	4	5	4	4	5	6	4	4	2	3	3	5
65	4	5	5	4	6	4	3	5	5	5	5	4	5	4	5
66	4	5	5	4	6	6	6	6	7	5	5	4	3	3	5
67	4	5	4	4	5	4	4	5	6	8	7	7	7	4	6
68	5	7	6	5	5	5	4	6	6	5	4	4	3	3	5
69	5	6	5	5	5	3	3	3	5	6	5	4	3	4	6
70	4	4	4	3	4	3	3	4	5	6	5	4	4	3	5
71	7	8	7	6	6	5	4	6	6	8	5	6	6	5	7
72	5	5	5	4	4	5	4	5	6	8	7	6	7	4	6
73	5	6	6	5	5	4	3	5	6	6	5	5	5	4	6
74	3	4	4	3	4	4	3	5	5	6	6	5	5	2	5
75	4	5	4	4	3	5	4	5	6	6	6	5	5	6	5
76	5	6	7	4	5	4	5	5	5	6	6	5	4	4	7
77	4	5	6	4	5	7	7	7	8	6	6	6	5	4	6
78	6	6	6	4	6	5	5	5	7	7	5	6	5	2	5
79	7	7	7	5	6	4	4	5	6	6	6	5	6	3	6
80	4	5	5	4	4	3	4	4	5	5	5	4	4	3	5
81	5	6	5	4	5	6	5	6	7	6	5	6	5	5	7
82	5	6	5	4	5	3	4	4	5	6	6	5	6	4	7
83	4	5	5	3	4	3	4	4	5	4	3	3	3	3	4
84	4	5	4	3	5	5	5	6	7	8	7	6	8	4	7
85	4	5	3	3	4	3	4	4	5	3	4	2	3	2	4
86	5	6	6	5	6	3	4	5	5	5	5	3	3	3	5
87	5	6	5	5	4	3	3	3	4	5	6	5	4	5	6
88	4	5	5	4	5	6	6	6	7	5	4	4	4	4	6
89	3	4	4	3	3	4	4	4	6	6	5	6	5	2	4
90	3	4	3	3	3	2	2	3	3	6	5	5	5	3	4
91	4	5	4	4	4	4	4	4	6	8	5	6	5	2	5
92	4	5	6	4	5	4	4	5	5	6	5	4	5	3	4
93	4	6	5	5	4	3	3	3	5	6	5	5	4	2	4
94	4	5	5	4	4	2	2	3	4	5	3	4	3	4	5

```
 95   4   5   5   3   4   3   4   3   5   6   5   5   5   4   6
 96   4   5   4   3   3   4   4   4   5   4   4   4   4   4   4
 97   4   5   4   4   5   3   4   4   5   5   5   4   4   1   6
 98   6   7   6   5   6   5   4   4   6   6   5   5   5   5   6
 99   7   8   6   6   6   3   4   4   4   7   6   5   5   4   6
100   6   7   7   5   6   5   5   6   7   6   6   5   5   6   7
101   4   5   4   4   4   4   3   3   6   4   4   4   5   3   5
102   4   5   5   4   3   3   4   4   5   5   5   4   4   3   4
103   7   8   7   6   7   4   5   5   5   7   8   7   6   4   7
104   2   2   3   2   2   4   5   4   6   6   5   5   6   3   4
105   4   5   4   3   5   4   4   4   6   7   8   6   6   6   6
106   5   6   6   5   6   3   3   5   4   6   6   4   5   3   4
107   4   5   4   3   5   5   4   5   5   6   5   4   5   2   6
108   4   5   4   3   3   4   3   5   6   5   5   4   4   4   6
109   7   7   7   6   6   4   4   4   5   6   6   5   4   4   6
110   4   6   6   4   4   2   3   3   4   5   7   5   5   4   5
111   5   6   7   5   4   5   5   4   6   6   5   5   5   4   6
112   5   6   6   5   4   5   5   4   6   4   3   3   4   6   6
113   5   6   4   4   6   5   4   5   6   6   5   6   5   5   6
114   4   5   6   3   4   3   3   4   4   5   5   4   4   4   5
115   3   3   3   2   3   2   3   3   4   6   4   3   4   2   5
116   5   6   6   5   6   4   4   5   5   5   4   3   3   3   6
117   6   6   6   5   6   4   5   4   5   7   7   6   4   3   7
118   3   3   3   1   3   4   3   4   5   6   5   5   4   4   5
119   5   6   6   5   6   2   3   3   4   5   4   4   5   2   4
120   6   7   5   5   6   4   4   5   6   4   4   3   4   5   5
121   3   4   4   2   3   3   3   4   5   5   4   5   4   2   4
;
run;
```

Untuk kasus nyata dalam melakukan analisis data tersebut, diperlukan tahapan eksplorasi secara lengkap termasuk pembuatan plot, korelasi dan statistika deskriptif. Namun pada modul ini tahapan tersebut dipersingkat dengan hanya menampilkan korelasinya saja.

```
proc corr data=contoh nosimple noprob;
       var hurtchew--facepain;
run;

                             The CORR Procedure

14 Variables:  hurtchew  hurtwide  noise  stiffjaw  cracking
               painslp   painrout paintab amhdache   earpain   grind
               clampset  sorejaw facepain

           hurtchew    hurtwide       noise    stiffjaw    cracking
hurtchew    1.00000     0.91857     0.73954     0.83351     0.74672
 hurtwide   0.91857     1.00000     0.75415     0.84741     0.70225
 noise      0.73954     0.75415     1.00000     0.74195     0.59856
 stiffjaw   0.83351     0.84741     0.74195     1.00000     0.67549
 cracking   0.74672     0.70225     0.59856     0.67549     1.00000
 painslp    0.06011     0.04377     0.09455    -0.02320     0.07515
 painrout   0.04807     0.00032     0.07731    -0.05749     0.07234
 paintab    0.10696     0.08191     0.11993     0.01156     0.18199
 amhdache   0.01362    -0.00149     0.06352    -0.06753     0.03717
 earpain    0.14590     0.09182     0.09979     0.10905     0.15655
 grind      0.19999     0.15790     0.16047     0.14768     0.22304
 clampset   0.18743     0.13108     0.14257     0.10542     0.12748

            painslp    painrout     paintab    amhdache     earpain
 hurtchew   0.06011     0.04807     0.10696     0.01362     0.14590
 hurtwide   0.04377     0.00032     0.08191    -0.00149     0.09182
 noise      0.09455     0.07731     0.11993     0.06352     0.09979
 stiffjaw  -0.02320    -0.05749     0.01156    -0.06753     0.10905
 cracking   0.07515     0.07234     0.18199     0.03717     0.15655
 painslp    1.00000     0.75940     0.78605     0.86336     0.16670
 painrout   0.75940     1.00000     0.68056     0.74152     0.12608
 paintab    0.78605     0.68056     1.00000     0.75483     0.14265
 amhdache   0.86336     0.74152     0.75483     1.00000     0.13100
 earpain    0.16670     0.12608     0.14265     0.13100     1.00000
 grind      0.06644     0.14109     0.08491     0.04135     0.72987
 clampset   0.19254     0.13345     0.13146     0.12323     0.76628

             grind       clampset       sorejaw       facepain
hurtchew    0.19999      0.18743       0.09069       0.34282
hurtwide    0.15790      0.13108       0.03075       0.35609
noise       0.16047      0.14257       0.01782       0.27994
stiffjaw    0.14768      0.10542       0.04959       0.22439
cracking    0.22304      0.12748       0.12789       0.17662
painslp     0.06644      0.19254       0.15206       0.34902
painrout    0.14109      0.13345       0.10004       0.19799
paintab     0.08491      0.13146       0.16825       0.29645
amhdache    0.04135      0.12323       0.15453       0.28534
earpain     0.72987      0.76628       0.76465       0.09168
grind       1.00000      0.69432       0.67947       0.14377
clampset    0.69432      1.00000       0.65412       0.17727
```

Bila Anda perhatikan hasil korelasi tersebut akan didapati beberapa peubah yang memiliki korelasi yang sangat besar. Hal itu terlihat jelas karena peubah-peubah tersebut diatur sedemikian rupa sehingga variabel yang berkorelasi cukup besar terletak berdampingan.

Dengan demikian tidak salah jika dari peubah-peubah yang berkorelasi tinggi tersebut, ada peubah-peubah laten atau faktor yang

mendasarinya. Analisis faktor akan digunakan untuk mendapatkan faktor yang dimaksud.

Untuk kasus di atas, perintah yang digunakan adalah

```
proc factor data=contoh method=ML rotate=VARIMAX nocorr
scree ;

      var hurtchew--facepain ;

      priors SMC;

run;
```

Penjelasan dari opsi-opsi yang dipakai pada contoh program di atas :

METHOD = ML	menyebutkan penggunaan metode kemungkinan maksimum dalam ekstraksi
NOCORR	tidak menampilkan matriks korelasi antar peubah teramati
SCREE	menampilkan scree plot, yaitu plot yang menjelaskan hubungan antara nilai akar ciri dengan banyaknya faktor
ROTATE = VARIMAX	menyebutkan metode rotasi faktor menggunakan VARIMAX
PRIORS = SMC	menyebutkan metode menentukan komunal menggunakan SMC

Dalam program di atas, kita tidak menyebutkan berapa faktor yang digunakan, sehingga SAS secara default akan menentukan, dan dari hasil didapatkan sebanyak 3 faktor.

Dugaan awal komunalitas dilakukan dengan metode SMC antar peubah teramati, hasilnya adalah sebagai berikut.

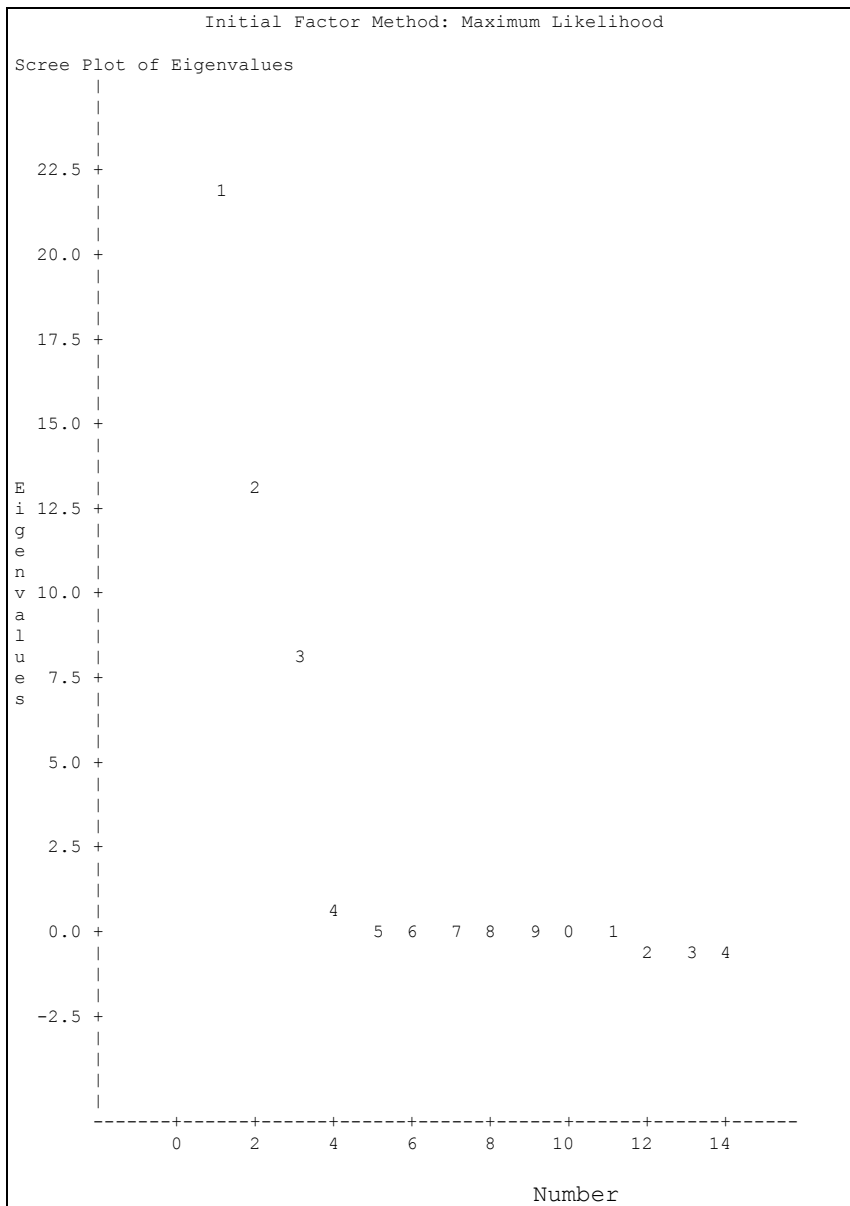
```
                        The FACTOR Procedure
               Initial Factor Method: Maximum Likelihood

                   Prior Communality Estimates: SMC

hurtchew    hurtwide       noise      stiffjaw      cracking
0.87838943 0.87757657  0.62831100     0.77519365     0.61159069

painslp     painrout       paintab      amhdache      earpain
0.82284339 0.65700913    0.67282808    0.78178793     0.73791420

           grind        clampset         sorejaw        facepain
      0.65023704      0.65812762      0.65210136      0.30773023
```

Selanjutnya, Anda akan mendapati akar ciri dari matriks korelasi yang tereduksi. Nilai-nilai tersebut diinterpretasikan sebagai ragam dari faktor yang berpadanan dengannya. Beberapa diantaranya bernilai negatif, yang merupakan ragam dari faktor yang imaginer.

```
     Preliminary Eigenvalues: Total = 42.2227014  Average = 3.01590724

              Eigenvalue    Difference    Proportion    Cumulative

         1    22.1866011     8.8566626        0.5255        0.5255
         2    13.3299385     5.4335408        0.3157        0.8412
         3     7.8963977     7.3961789        0.1870        1.0282
         4     0.5002189     0.2362913        0.0118        1.0400
         5     0.2639276     0.0650434        0.0063        1.0463
         6     0.1988841     0.1909592        0.0047        1.0510
         7     0.0079249     0.0792444        0.0002        1.0512
         8    -0.0713195     0.1620991       -0.0017        1.0495
         9    -0.2334186     0.0273054       -0.0055        1.0440
        10    -0.2607240     0.0411412       -0.0062        1.0378
        11    -0.3018652     0.0570478       -0.0071        1.0306
        12    -0.3589130     0.0828409       -0.0085        1.0221
        13    -0.4417539     0.0514436       -0.0105        1.0117
        14    -0.4931974                    -0.0117        1.0000

         3 factors will be retained by the PROPORTION criterion.
```

Kriteria default jika menggunakan METHOD=ML adalah mengambil faktor sehingga persentase keragaman yang dijelaskan mencapai 100%. Dalam contoh di atas, terlihat bahwa dengan menggunakan 3 faktor telah tercapai kumulatif sebesar 1.00.

Dengan berdasarkan scree plot juga dapat dilihat bahwa tiga akar ciri pertama jauh lebh besar daripada akar ciri sisanya, yang menerangkan bahwa faktor sisanya memeilki kontribusi yang jauh lebih kecil.

```
                        The FACTOR Procedure
```

Metode kemungkinan maksimum menggunakan prosedur optimasi non-linear secara iteratif. Pada contoh ini diperlukan tiga iterasi untuk mencapi kekonvergenan.

```
                         The SAS System

                       The FACTOR Procedure
            Initial Factor Method: Maximum Likelihood

Iteration    Criterion    Ridge   Change     Communalities

    1        0.4720449   0.0000   0.0967   0.91413   0.91818   0.63165
                                           0.78954   0.57923   0.89442
                                           0.64866   0.69852   0.83648
                                           0.83458   0.68004   0.68794
                                           0.68317   0.21737
    2        0.4712085   0.0000   0.0081   0.91491   0.92098   0.62716
                                           0.78520   0.57855   0.89474
                                           0.64929   0.69613   0.83615
                                           0.83603   0.67300   0.69218
                                           0.68715   0.22544
    3        0.4711972   0.0000   0.0009   0.91533   0.92095   0.62668
                                           0.78494   0.57798   0.89481
                                           0.64923   0.69610   0.83611
                                           0.83692   0.67288   0.69165
                                           0.68669   0.22587

   Convergence criterion satisfied.
```

Terdapat uji awal berupa uji nisbah kemungkinan untuk mengevaluasi seberapa bagus jumlah faktor sebayak tiga yang terpilih tadi. Dalam hal ini bisa dilakukan pengujian terhadap hipotesis nol tentang tidak ada faktor, atau bisa juga diuji hipotesis nol yang berupa tidak dibutuhkan lagi lebih banyak faktor. Pada contoh ini, kita tolah hipotesis nol yang pertama, dan kita terima hipotesis nol yang kedua.

```
                 Significance Tests Based on 121 Observations

                                                            Pr >
                Test                       DF    Chi-Square    ChiSq

    H0: No common factors                  91     1356.7888    <.0001
    HA: At least one common factor
    H0: 3 Factors are sufficient           52       53.0097    0.4350
    HA: More factors are needed

        Chi-Square without Bartlett's Correction        56.54367
        Akaike's Information Criterion                 -47.45633
        Schwarz's Bayesian Criterion                  -192.83744
        Tucker and Lewis's Reliability Coefficient       0.99860
```

Berikut adalah ouput dari analisis faktor yang belum dirotasi. Namun, untuk interpretasi akan digunakan hasil setelah dirotasi.

```
                    Squared Canonical Correlations

                Factor1         Factor2         Factor3
             0.96766451      0.94865487      0.91245558

            Eigenvalues of the Weighted Reduced Correlation
```

```
          Matrix: Total = 58.8245826  Average = 4.2017559

           Eigenvalue    Difference    Proportion    Cumulative
      1    29.9257685    11.4497265        0.5087        0.5087
      2    18.4760420     8.0532689        0.3141        0.8228
      3    10.4227732     9.9002273        0.1772        1.0000
      4     0.5225458     0.1755004        0.0089        1.0089
      5     0.3470454     0.0329634        0.0059        1.0148
      6     0.3140820     0.1781226        0.0053        1.0201
      7     0.1359594     0.0777242        0.0023        1.0224
      8     0.0582352     0.1223935        0.0010        1.0234
      9    -0.0641583     0.0469919       -0.0011        1.0223
     10    -0.1111502     0.0885464       -0.0019        1.0204
     11    -0.1996966     0.0414323       -0.0034        1.0170
     12    -0.2411290     0.0959318       -0.0041        1.0129
     13    -0.3370608     0.0876132       -0.0057        1.0072
     14    -0.4246740                   -0.0072        1.0000

                        Factor Pattern
                    Factor1         Factor2         Factor3
      hurtchew      0.94910        -0.11071        -0.04773
      hurtwide      0.94297        -0.14657        -0.10148
      noise         0.78511        -0.05376        -0.08561
      stiffjaw      0.86188        -0.19845        -0.05177
      cracking      0.75905        -0.04229        -0.01028
      painslp       0.16230         0.90769        -0.21114
      painrout      0.11640         0.77729        -0.17748
      paintab       0.19416         0.78821        -0.19268
      amhdache      0.10793         0.87939        -0.22613
      earpain       0.23652         0.32783         0.82067
      grind         0.27884         0.21407         0.74109
      clampset      0.26239         0.31251         0.72470
      sorejaw       0.16759         0.32315         0.74447
      facepain      0.38357         0.27465        -0.05768

                  Variance Explained by Each Factor

                 Factor        Weighted    Unweighted
                 Factor1     29.9257685    4.19235839
                 Factor2     18.4760420    3.33128063
                 Factor3     10.4227732    2.49250606

Final Communality Estimates and Variable Weights
   Total Communality: Weighted = 58.824584   Unweighted = 10.016145

                Variable    Communality        Weight

                hurtchew     0.91532762    11.8099537
                hurtwide     0.92096839    12.6510054
                noise        0.62661878     2.6786820
                stiffjaw     0.78489781     4.6498962
                cracking     0.57804504     2.3695735
                painslp      0.89481246     9.5067471
                painrout     0.64922382     2.8508558
                paintab      0.69609785     3.2905561
                amhdache     0.83610851     6.1016255
                earpain      0.83691756     6.1321039
                grind        0.67278898     3.0569829
                clampset     0.69170415     3.2430928
                sorejaw      0.68674643     3.1917282
                facepain     0.22588769     1.2917796
```

Berikut adalah hasil rotasi. Yang ingin dievaluasi adalah ROTATED FACTOR PATTERN. Secara umum, ada dua jenis matriks yang menggambarkan hubungan antara peubah laten (faktor) dan peubah teramati :

- matriks FACTOR PATTERN, berisi koefisien untuk menyusun peuabh teramati dari peubah faktor
- matriks FACTOR STRUCTURE, berisi korelasi antara faktor dengan peubah teramati.

Setelah mengalami rotasi, kedua matriks diatas identik.

Rotated Factor Pattern

	Factor1	Factor2	Factor3
hurtchew	0.95158	0.03084	0.09422
hurtwide	0.95904	0.01259	0.03239
noise	0.78672	0.07526	0.04504
stiffjaw	0.88236	-0.06184	0.05018
cracking	0.74922	0.05982	0.11462
painslp	0.02799	0.94176	0.08435
painrout	0.00150	0.80258	0.07133
paintab	0.07747	0.82742	0.07400
amhdache	-0.01810	0.91270	0.05254
earpain	0.06377	0.08949	0.90821
grind	0.13556	0.01174	0.80887
clampset	0.10451	0.10757	0.81805
sorejaw	0.00754	0.09952	0.82267
facepain	0.33392	0.32564	0.09137

Penjelasan dari setiap faktor adalah sebagai berikut:

- Faktor 1, menerangkan peubah-peubah yang berhubungan dengan pergerakan rahang
- Faktor 2, menerangkan peubah-peubah yang berhubungan dengan keluhan-keluhan yang diderita
- Faktor 3, menerangkan tentang rasa sakit di gigi dan rahang

Peubah terakhir, yaitu FACEPAIN berhubungan cukup erat dengan dua faktor pertama.

Variance Explained by Each Factor

Factor	Weighted	Unweighted
Factor1	29.2279839	3.93610681
Factor2	17.9138690	3.19852482
Factor3	11.6827308	2.88151346

Final Communality Estimates and Variable Weights
Total Communality: Weighted = 58.824584 Unweighted = 10.016145

Variable	Communality	Weight

```
hurtchew     0.91532762    11.8099537
hurtwide     0.92096839    12.6510054
noise        0.62661878     2.6786820
stiffjaw     0.78489781     4.6498962
cracking     0.57804504     2.3695735
painslp      0.89481246     9.5067471
painrout     0.64922382     2.8508558
paintab      0.69609785     3.2905561
amhdache     0.83610851     6.1016255
earpain      0.83691756     6.1321039
grind        0.67278898     3.0569829
clampset     0.69170415     3.2430928
sorejaw      0.68674643     3.1917282
facepain     0.22588769      1.2917796
```

8.9. Latihan

1. Berikut adalah data panjang tulang ramus pada empat kelompok umur dari 20 anak laki-laki.

	Umur (Tahun)			
Individu	8 (y_1)	$8\frac{1}{2}$ (y_2)	9 (y_3)	$9\frac{1}{2}$ (y_4)
1	47.8	48.8	49.0	49.7
2	46.4	47.3	47.7	48.4
3	46.3	46.8	47.8	48.5
4	45.1	45.3	46.1	47.2
5	47.6	48.5	48.9	49.3
6	52.5	53.2	53.3	53.7
7	51.2	53.0	54.3	54.5
8	49.8	50.0	50.3	52.7
9	48.1	50.8	52.3	54.4
10	45.0	47.0	47.3	48.3
11	51.2	51.4	51.6	51.9
12	48.5	49.2	53.0	55.5
13	52.1	52.8	53.7	55.0
14	48.2	48.9	49.3	49.8
15	49.6	50.4	51.2	51.8
16	50.7	51.7	52.7	53.3
17	47.2	47.7	48.4	49.5
18	53.3	54.6	55.1	55.3
19	46.2	47.5	48.1	48.4
20	46.3	47.6	51.3	51.8

Lakukan analisis faktor dengan metode komponen utama.

9

9.alisis Gerombol (*Cluster Analysis*)

9.1. Pendahuluan

Analisis gerombol merupakan teknik peubah ganda yang mempunyai tujuan utama untuk mengelompokkan objek-objek berdasarkan kemiripan karakteristik yang dimilikinya. Karakteristik objek-objek dalam suatu gerombol memiliki tingkat kemiripan yang tinggi, sedangkan karakteristik antar objek pada suatu gerombol dengan gerombol lain memiliki tingkat kemiripan yang rendah. Dengan kata lain, keragaman dalam suatu gerombol minimum sedangkan antar keragaman antar gerombol maksimum.

Variabel-variabel yang dilibatkan dalam analisis gerombol dipilih sesuai dengan tujuan penggerombolan. Misalnya dalam suatu kajian ingin dibentuk gerombol objek berdasarkan kemampuan prestasi akademik. Variabel-variabel yang relevan dimasukan dalam kajian ini antara lain nilai rapor setiap mata pelajaran yang diambil dan nilai ujian nasional. Berbeda dengan teknik multivariat lainnya, analisis ini tidak memilih gugus variabel secara empiris tetapi menggunakan gugus variabel yang ditentukan oleh peneliti itu sendiri sesuai dengan tujuan penggerombolan.

Kemiripan antar objek diukur dengan menggunakan ukuran jarak. Beberapa ukuran jarak yang sering digunakan antara lain jarak Euclid, jarak mahalanobis, jarak City-block (Manhattan), dan lain-lain.

1. **Jarak Euclidean**

Ini merupakan tipe jarak yang paling umum dipilih. Kemudahannya adalah jarak geometrik dalam ruang dimensi ganda. Perhitungannya sebagai berikut:

$$\text{Jarak } (x,y)=\{i(xi-yi)2\}1/2 .$$

Perhatikan bahwa jarak Euclidean (dan kuadrat Euclidean) biasanya dihitung dari data mentah, dan tidak dari data standar. Metode ini memiliki beberapa keuntungan, antara lain jarak dari 2 objek apa saja tidak dipengaruhi oleh penambahan dari objek baru untuk di analisis, yang mungkin merupakan pencilan. Namun demikian, jarak bisa menjadi sangat besar, disebabkan hanya karena perbedaan skala. Sebagai contoh, jika suatu dimensi di ukur dalam satuan jarak cm, dan dikonversi dalam mm (dengan mengkalikan nilai dengan 10), hasil dari jarak Euklidean atau kuadrat Euklidean (dihitung dari dimensi ganda) bisa sangat berbeda, dan konsekuensinya, hasil dari analisis kluster mungkin menjadi berbeda.

2. **Jarak Kuadrat Euclidean**

Mungkin kita ingin mengkuadratkan jarak Euclidean standar untuk menempatkan bobot yang lebih besar secara progresif pada objek yang jaraknya jauh. Jarak ini perhitungannya sbb:

$$\text{jarak } (x.y)=\Sigma_i(x_i-y_i)^2 .$$

3. **Jarak City-Block (Manhattan**)

Jarak ini memudahkan jarak rata-rata dimensi-dimensi secara menyilang. Dalam banyak kasus, ukuran jarak ini mendapatkan hasil yang mirip terhadap jarak Euklidean sederhana. Bagaimanapun, perhatikan bahwa dalam ukuran ini, efek dari perbedaan besar yang tunggal (pencilan) dibuang, karena tidak dikuadratkan.

arak City-Block ditulis sbb:

$$\text{jarak}(x,y) = \Sigma i \mid xi - yi \mid$$

4. **Jarak Chebychev**

   Ukuran jarak ini mungkin cocok untuk kasus menentukan 2 objek, sebagai "Berbeda" jika mereka berbeda terhadap segala sesuatu pada dimensi. Jarak Chebychev dihitung sbb :

$$\text{jarak}(x,y) = \text{maksimum} \mid xi - yi \mid$$

5. **Jarak Kuasa**

   Untuk meningkatkan atau menurunkan bobot progresiv yang ditempatkan pada dimensi yang respektif terhadap objek yang berbeda bisa dipenuhi oleh jarak kuasa. Jarak kuasa dihitung sbb:

$$\text{jarak}(x,y) = (\Sigma_i \mid X_i - Y_i \mid^p)^{1/r}$$

   Dimana r dan p adalah parameter yang ditentukan. Beberapa contoh penghitungan akan didemonstrasikan, bagaimana ini mengukur "perilaku". Parameter p mengontrol bobot progresiv yang ditempatkan pada perbedaan di dimensi individu, parameter r mengontrol bobot progresiv yang ditempatkan pada perbedaan yang lebih besar antara objek . jika r dan p bernilai 2, maka jarak ini sama dengan jarak Euklidean.

9.2. Metode Penggerombolan

Terdapat tiga metode yang umum digunakan dalam penggerombolan objek yaitu metode grafik, metode berhirarki dan metode tak berhirarki.

9.2.1. Metode Grafik

Metode grafik, terdiri atas tiga jenis yaitu plot profil, plot Andrew dan plot Andrew termodifikasi. Pendekatan grafik yang paling sederhana adalah

menggunakan plot profil dari setiap pengamatan. Plot ini hanya efektif untuk data yang tidak terlalu banyak pengamatannya, sehingga pembakuan data sangat membantu dalam proses ini.

Plot lain yang bias digunakan adalah plot Andrew dan plot Andrew termodifikasi. Kedua plot ini memberikan hasil yang lebih efektif dalam penggerombolan objek. Untuk vector pengamatan $x = (x_1, x_2, \ldots x_p)'$, fungsi Andrew dan dan fungsi Andrew termodifikasi berturut – turut didefinisikan sebagai berikut:

$$f_x t = \frac{1}{\sqrt{2}} x_1 + x_2 \sin(t) + x_3 \cos(t) + x_4 \sin(2t) + x_5 \cos(2t) + \cdots, -\pi \leq t \leq \pi$$

dan

$$g_x t = \frac{1}{\sqrt{2}} [x_1 + x_2[\sin(t) + \cos(t)] + x_3 [\sin(t) - \cos(t)] + x_4 [\sin(2t) + \cos(2t)] + x_5 [\sin(2t) - \cos(2t)] + \cdots], -\pi \leq t \leq \pi$$

Ketiga metode ini sama – sama menggambarkan karakteristik tiap gerombol. Titik bertanya pada karakteristik yang secara signifikan berbeda antar gerombol dan memprediksi anggota dalam suatu gerombol khusus.

9.2.2. Metode Penggerombolan Berhierarkhi

Metode ini digunakan untuk menggerombolkan pengamatan secara terstuktur berdasarkan kemiripan sifatnya dan gerombol yang diinginkan belum diketahui banyaknya. Ada dua cara untuk mendapatkan gerombol dengan metode penggerombolan hirarki yaitu dengan cara penggabungan (*aglomerative*) dan pemisahan gerombol (*devisive*). Metode hirarki dengan cara penggabungan didapat dengan menggabungkan pengamatan atau gerombol secara bertahap, sehingga pada akhirnya didapat hanya satu gerombol saja. Sebaliknya

cara pemisahan pada metode hirarki dimulai dengan membentuk satu gerombol besar beranggotakan seluruh pengamatan. Gerombol besar tersebut kemudian dipisah menjadi gerombol yang lebih kecil, sampai satu gerombol hanya beranggotakan satu pengamatan saja. Kedua cara dalam metode hirarki ini tidak berbeda dalam pembentukan gerombol yang terjadi tetapi hanya berbeda dalam tahapan pembentukan gerombol saja.

9.2.2.1. *Keterhubungan Tunggal (Single Linkage/Nearest Neighbor)*

Metode *Single Linkage* atau disebut juga dengan motode pautan tunggal, jarak antara dua cluster (kelompok) dapat ditentukan dari dua obyek berpasangan yang memiliki kesamaan atau memiliki jarak terdekat (*nearest neighbor*) dalam cluster yang berbeda. Hal ini berlaku jika pengelompokan itu mempunyai makna atau tujuan yang jelas, terdapatnya kumpulan-kumpulan obyek yang bersama-sama membentuk cluster serta kecenderungan dari hasil pengelompokan itu menggambarkan rantai "chain" yang panjang.

Pertama kali yang harus dilakukan untuk mendapatkan pautan tunggal ini adalah harus menemukan jarak terdekat antar cluster-cluster tersebut, dirumuskan dengan $D = \{d_{ik}\}$. Kemudian menggabungkan obyek-obyek yang sesuai, misalkan obyek tersebut dilambangkan dengan cluster U dan cluster V untuk mendapatkan cluster gabungan (UV).

Untuk menghitung jarak cluster (UV) dengan cluster-cluster yang lain dapat dirumuskan dengan :

$$d_{(UV)W} = \min \{d_{UW}, d_{VW}\}$$

<u>Catatan:</u> nilai d_{UW} dan d_{VW} menggambarkan jarak terdekat antara cluster U dengan W serta V dengan W. Hasil dari single linkage clustering ini

dapat ditampilkan dalam bentuk dendogram atau diagram pohon, dimana dahan atau cabang dari diagram pohon tersebut menggambarkan clusternya.

9.2.2.2. *Keterhubungan menyeluruh (Complete Linkage / Furthest Neighbor)*

Jarak antar cluster pada metode *Complete Linkage* atau disebut juga metode pautan lengkap, ditentukan dari jarak terjauh antara dua obyek pada cluster yang berbeda (*furthest neighbor*). Metode ini dapat digunakan dengan baik untuk kasus dimana obyek-obyek yang ada berasal dari kelompok yang benar-benar berbeda.

Langkah pertama yang harus dilakukan adalah sama seperti kasus *single linkage* yaitu harus menemukan jarak terdekat antar cluster-cluster tersebut, dirumuskan dengan $D = \{d_{ik}\}$. Kemudian menggabungkan obyek-obyek yang sesuai, misalkan obyek tersebut dilambangkan dengan cluster U dan cluster V untuk mendapatkan cluster gabungan (UV).

Untuk menghitung jarak cluster (UV) dengan cluster-cluster yang lain dapat dirumuskan dengan :

$$d_{(UV)W} = \max \{d_{UW}, d_{VW}\}$$

<u>Catatan:</u> nilai d_{UW} dan d_{VW} menggambarkan jarak terjauh antara cluster U dengan W serta V dengan W.

9.2.2.3. *Keterhubungan Rataan (Average Linkage)*

Jarak antar cluster pada metode *Average Linkage* atau disebut juga metode pautan rataan, ditentukan dari rata-rata jarak seluruh objek suatu cluster terhadap seluruh objek pada cluster lainnya. Pada berbagai keadaan, metode ini dianggap lebih stabil dibandingkan dengan kedua metode di atas.

Langkah pertama yang harus dilakukan adalah sama seperti kasus *single atau complete linkage* yaitu harus menemukan jarak terdekat antar cluster-cluster tersebut, dirumuskan dengan $D = \{d_{ik}\}$. Kemudian menggabungkan obyek-obyek yang sesuai, misalkan obyek tersebut dilambangkan dengan cluster U dan cluster V untuk mendapatkan cluster gabungan (UV).

Untuk menghitung jarak cluster (UV) dengan cluster-cluster yang lain dapat dirumuskan dengan :

$$d_{(UV)W} = \text{rata-rata } \{d_{UW}, d_{VW}\}$$

<u>Catatan:</u> nilai d_{UW} dan d_{VW} menggambarkan jarak rata-rata antara cluster U dengan W serta V dengan W.

9.2.2.4. *Algoritma dan metode pengelompokan hirarki lainnya*

Secara umum, algoritma untuk metode pengelompokan hirarki adalah :

1. Tentukan matriks jarak antar data atau kelompok
2. Gabungkan dua data atau kelompok terdekat ke dalam kelompok yang baru
3. Tentukan kembali matrik jarak tersebut.
4. Lakukan langkah 2 dan 3 sampai semua data masuk dalam satu kelompok

Hasil pengelompokan tersebut dapat digambarkan dengan diagram pohon 3 dendrogram atau *vicicle plot.* Jumlah kelompok yang terjadi ditentukan dari dendrogram yang terjadi dan tergantung subyektivitas peneliti. Namun demikian pemisahan kelompok biasanya ditentukan berdasarkan jarak penggabungan terbesar.

Sebenarnya masih banyak jenis dari metode analisis hierarki, *Unweighted Pair-Group Average, Weighted Pair-Group Average, Unweighted Pair-*

Group Centroid, Weighted Pair-Group Centroid, dan *Ward's Method***,** tetapi karena kompleksitas pemakaiannya, motode ini relatif jarang digunakan.

9.2.2.5. *Ilustrasi Metode Hierarkhi*

Kasus 1:

Salah satu contoh sederhana dari *single lingkage* akan diuraikan berikut ini.

Misalkan kita memiliki matriks jarak D sebagai berikut:

$D = \{d_{ik}\} =$

	1	2	3	4	5
1	0				
2	9	0			
3	3	7	0		
4	6	5	9	0	
5	11	10	2	8	0

Tahapan:

1. Clustering dilakukan dengan menggabungkan dua obyek yang memiliki jarak terdekat, yaitu :

$$\min_{i,k} (d_{ik}) = d_{53} = 2$$

2. Gabungkan Obyek 5 dan 3 menjadi satu cluster (35) dan selanjutnya hitung jarak objek lainnya terhadap cluster (35), seperti berikut:

$$d_{(35)1} = \min\{d_{31}, d_{51}\} = \min\{3,11\} = 3$$

$$d_{(35)2} = \min\{d_{32}, d_{52}\} = \min\{7,10\} = 7$$

$$d_{(35)4} = \min\{d_{34}, d_{54}\} = \min\{9,8\} = 8$$

Sehingga pengelompokan yang baru diperoleh :

	(35)	1	2	4
(35)	0			
1	3	0		
2	7	9	0	
4	8	6	5	0

3. Pengelompokan berikutnya yaitu penggabungan cluster 1 dengan (35) karena merupakan jarak terdekat, dan kemudian menghitung kembali jarak cluster lain terhadap cluster (135) tersebut:

   $d_{(135)2} = \min \{d_{(35)2}, d_{12}\} = \min \{7,9\} = 7$

   $d_{(135)4} = \min \{d_{(35)4}, d_{14}\} = \min \{8,6\} = 6$

   Sehingga pengelompokan yang baru diperoleh :

	(135)	2	4
(135)	0		
2	7	0	
4	6	5	0

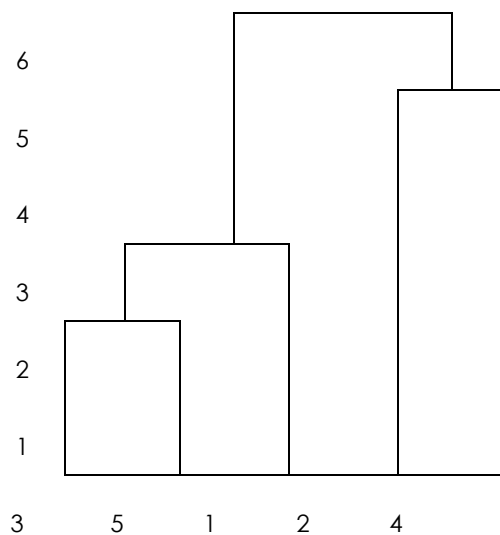
4. Pengelompokan berikutnya yaitu penggabungan cluster 2 dengan 4 karena merupakan jarak terdekat, dan kemudian menghitung kembali jarak cluster lain terhadap cluster (24) tersebut:

   $d_{(135)(24)} = \min \{d_{(135)2}, d_{(135)4}\} = \min \{7,6\} = 6$

   Sehingga pengelompokan yang baru diperoleh :

	(135)	(24)
(135)	0	
(24)	6	0

5. Untuk selanjutnya cluster (135) dengan cluster (24) bergabung membentuk cluster tunggal dari kelima obyek tadi, sehingga diperoleh cluster (12345) dengan jarak terdekatnya = 6.

Hasil dari proses perhitungan ini biasanya ditampilkan dalam bentuk dendogram sebagai berikut:

Kasus 2:

Misalkan ingin dilakukan pengelompokan propinsi-propinsi di pulau Jawa berdasarkan karakteristik penduduk lanjut usia (lansia). Hal ini dilakukan mengingat penduduk lansia semakin lama semakin banyak (penduduk lansia cenderung bertambah atau ada perubahan struktur dari struktur

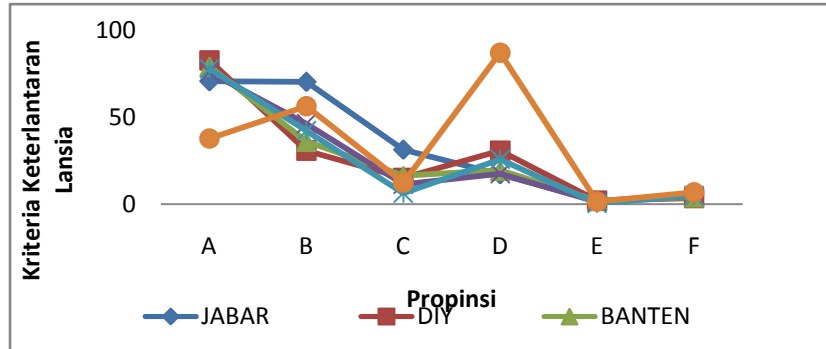
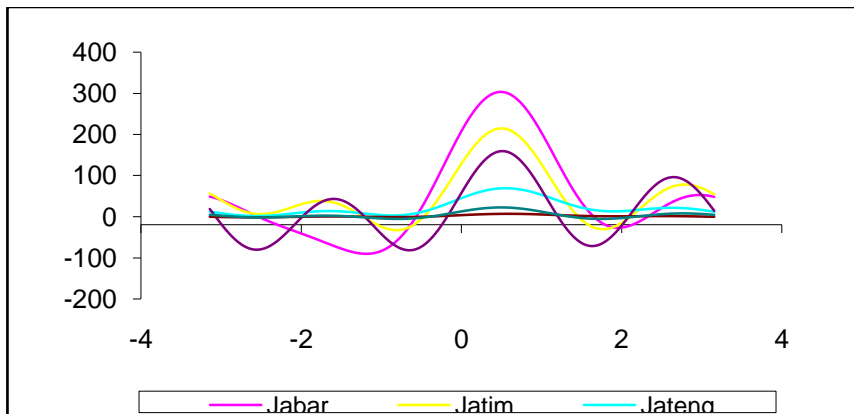
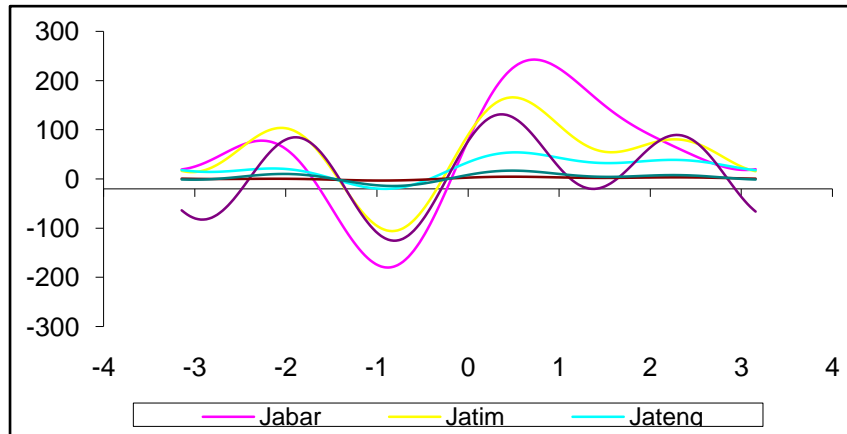
penduduk muda ke struktur penduduk tua), sehingga perlu adanya perhatian ekstra dari pihak pemerintah untuk permasalahn tersebut. Digunakan enam criteria dalam penelitian ini, yaitu tidak pernah sekolah/tamat SD (sebagai criteria A), makan makanan pokok <21x/minggu (sebagai criteria B), makan lauk pauk berprotein tinggi <4x/minggu (sebagai criteria C), memiliki pakaian <4 stel (sebagai criteria D), tidak mempunyai tempat tetap untuk tidur (sebagai criteria E) dan bila sakit tidak di obati (sebagai criteria F).

Tabel 9.1 Enam Kriteria Keterlantara Lansia Menurut Propinsi di pulau Jawa

PROPINSI	tidak pernah sekolah/tamat SD	makan makanan pokok <21X dalam seminggu	makan lauk pauk berprotein tinggi<4X dalam seminggu	memiliki pakaian <4 stel	tidak mempuny ai tempat tetap untuk tidur	bila sakit tidak diob ati
JABAR	70.84	70.48	31.37	17.17	1.82	5.32
JATIM	79.3	35.99	16.25	19.36	1.89	3.97
JATENG	79.3	35.99	16.25	19.36	1.89	3.97
DIY	76.05	46.27	11.35	17.45	1.17	4.72
BANTEN	77.96	42.48	6.28	25.74	0.34	4.87
DKI	37.8	56.38	12.28	87.24	1.45	6.78

Berikut akan dibahas beberapa metode yang digunakan dalam analisis gerombol

a. Metode grafik

Dalam metode ini menitikberatkan pada karakteristik yang secara signifikan berbeda antar gerombol dan memprediksi anggota dalam suatu gerombol khusus. Ada 3 plot yang digunakan yaitu plot profil, plot andrew dan plot Andrew termodifikasi.

Gambar 9.1 Plot Profil

Gambar 9.2 Plot Andrew

Gambar 9.3 Plot Andrew termodifikasi

Berdasarkan ketiga plot diatas, dapat disimpulkan bahwa untuk propinsi DIY, BANTEN, JATENG, JATIM mempunyai profil atau karakteristik yang mirip sedangkan JABAR dan DKI mempunyai karekteristik sendiri – sendiri.

Ukuran kemiripan dan ketakmiripan

Sebelum melakukan analisis lebih lanjut, untuk menentukan kemiripan antar objek dilihat berdasarkan kedekatan jarak antar objek tersebut. Makin dekat jaraknya, maka makin mirip karakteristik kedua objek tersebut. Jarak yang paling umum digunakan adalah *jarak Euclidean,* yang mengukur jarak sesungguhnya menggunakan mata manusia. Dengan menggunakan data pada table 1, jika dihitung jarak Euclidean akan didapatkan matriks jarak sebagai berikut

Tabel 9.2. Matriks jarak

	Jabar	Jatim	Jateng	DIY	Banten	DKI
Jabar	0	46.51556	38.68294	31.85796	39.3863	81.03674
Jatim	46.51556	0	13.03541	21.82378	16.1744	76.68378
Jateng	38.68294	13.03541	0	12.04076	13.58961	82.27726
DIY	31.85796	21.82378	12.04076	0	10.71021	80.2565
Banten	39.3863	16.1744	13.58961	10.71021	0	75.06508
DKI	81.03674	76.68378	82.27726	80.2565	75.06508	0

b. Metode Berhirarki

Metode hirarki paling banyak digunakan oleh para peneliti karena memiliki keunggulan tersendiri dimana pengelompokan yang terbentuk dapat terjadi secara alamiah. Awalnya dianggap n buah gerombol yang masing – masing beranggotakan satu objek. Dua gerombol yang paling dekat kemudian digabung dan selanjutnya ditentukan kembali kedekatan (atau jarak) antar (n-1) gerombol yang baru. Proses tersebut berlangsung sampai didapatkan hanya satu gerombol yang anggotanya adalah seluruh objek. Metode ini terdiri atas lima jenis yaitu:

1. Pautan tunggal (tetangga terdekat)

   Dari matriks jarak yang diperoleh (tabel 2), yang memiliki jarak terdekat yaitu (DIY, Banten). Dengan demikian kedua objek tersebut dapat dikelompokkan dalam satu gerombol. Gerombol yang dihasilkan tadi, kemudian dibandingkan dengan objek yang lain, yang memiliki jarak terdekat dengan gerombol yang terbentuk.

- h{(DIY, Banten), Jabar} = min {d(DIY, Jabar), d(Banten, Jabar)

  = min (31.85796, 39.3863)

  = 31.85796

- h{(DIY, Banten), Jatim} = min {d(DIY, Jatim), d(Banten, Jatim)}

  = min (21.82378, 16.1744)

  = 16.1744

- h{(DIY, Banten), Jateng} = min {d(DIY, Banten), d(Jateng, Banten)}

  = min (10.71021, 13.58961)

  = 10.71021

- h{{(DIY, Banten), DKI} = min { d(DIY, DKI) ,d(Banten, DKI)}

  = min (80.2565, 75.06508)

  = 75.06508

	Jabar	**Jatim**	**Jateng**	**(DIY, Banten)**	**DKI**
Jabar	0	46.51556	38.68294	31.85796	81.03674
Jatim	46.51556	0	13.03541	16.1744	76.68378
Jateng	38.68294	13.03541	0	10.71021	82.27726
(DIY, Banten)	31.85796	16.1744	10.71021	0	75.06508
DKI	81.03674	76.68378	82.27726	75.06508	0

Dari matriks jarak diatas diperoleh jarak antara (DIY, Banten) dan Jateng yang paling kecil, sehingga dijadikan sebagai satu gerombol. Jadi matriks jarak baru yang terbentuk sebagai berikut:

- h{(Jateng, DIY, Banten), Jabar}

  = min {d(Jateng, Jabar), d(DIY, Jabar), d(Banten, Jabar)}

  = min (38.68294, 31.85796, 39.3863)

  = 31.85796

- h{(Jateng, DIY, Banten), Jatim}

  = min {d(Jateng, Jatim), d(DIY, Jatim), d(Banten, Jatim)}

  = min (13.03541, 21.82378, 16.1744)

  = 13.03541

- h{(Jateng, DIY, Banten), DKI}

  = min {d(Jateng, DKI), d(DIY, DKI), d(Banten, DKI)}

  = min (82.27726, 80.2565, 75.06508)

  = 75.06508

	Jabar	**Jatim**	**(Jateng, DIY, Banten)**	**DKI**
Jabar	0	46.51556	31.85796	81.03674
Jatim	46.51556	0	13.03541	76.68378
(Jateng, DIY, Banten)	31.85796	13.03541	0	75.06508
DKI	81.03674	76.68378	80.2565	0

Dari matriks jarak diatas diperoleh jarak antara (Jateng, DIY, Banten) dan Jatim yang paling kecil, sehingga dijadikan sebagai satu gerombol. Jadi matriks jarak baru yang terbentuk sebagai berikut:

- h{(Jateng, DIY, Banten, Jatim), Jabar}

  = min{d(Jateng,Jabar), d(DIY,Jabar), d(Banten,Jabar), d(Jatim,Jabar)}

  = min(36.68294, 31.85796, 39.3863, 46,51556)

  = 31.85796

- h{(Jateng, DIY, Banten, Jatim), DKI}

  = min{d(Jateng, DKI), d(DIY, DKI), d(Banten, DKI), d(Jatim, DKI)}

  = min(82.27726, 80.2565, 75.06508, 76,68378)

  = 75.06508

	Jabar	**(Jateng, DIY, Banten, Jatim)**	**DKI**
Jabar	0	31.85796	81.03674
(Jateng, DIY, Banten, Jatim)	31.85796	0	75.06508
DKI	81.03674	75.06508	0

Dari matriks jarak baru yang terbentuk diperoleh jarak antara (Jateng, DIY, Banten, Jatim) dan Jabar yang paling kecil, sehingga dijadikan sebagai satu gerombol. Matriks jarak baru yang terbentuk sebagai berikut:

- h{(Jateng,DIY,Banten,Jatim,Jabar),DKI}

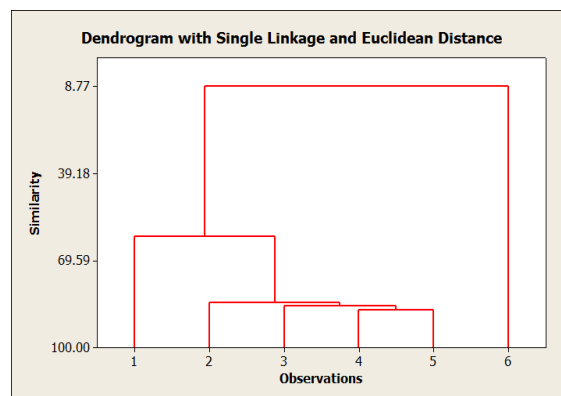
= min{d(Jateng,DKI), d(DIY,DKI), d(Banten,DKI), d(Jatim,DKI), (Jabar,DKI)}

= min(82.27726, 80.2565, 75.06508, 76,68378, 81.03674)

= 75.06508

	(Jateng, DIY, Banten, Jatim, Jabar)	**DKI**
(Jateng, DIY, Banten, Jatim, Jabar)	0	75.06508
DKI	75.06508	0

Dari matriks diatas akhirnya akan terbentuk satu gerombol yang terdiri dari seluruh propinsi di pulau Jawa yaitu (Jateng, DIY, Banten, Jatim, Jabar, dan DKI)

Dari iterasi diatas, maka dapat digambarkan dalam bentuk dendogram berikut:

Dari dendogram tersebut dapat dilihat bahwa pengklasteran dimulai dengan penggerombolan antara DIY dan Banten sebagai satu gerombol, pada iterasi yang kedua Jateng bergerombol dengan (DIY dan Banten), iterasi ketiga (Jateng, DIY dan Banten) bergerombol dengan Jatim, iterasi keempat (Jatim, Jateng, DIY dan Banten) bergerombol dengan Jabar, dan pada iterasi yang terakhir semua propinsi (Jabar, Jatim, Jateng, DIY, Banten, dan DKI) bergerombol menjadi satu gerombol. Jadi, jika kita menginginkan dua gerombol, maka gerombol pertama terdiri dari Jatim, Jateng, DIY dan Banten dan gerombol kedua terdiri dari Jabar dan DKI.

Algoritma penyelesaian dengan menggunakan pautan lengkap, rata-rata, ward, dan centroid pada dasarnya penyelesaiannya sama dengan pautan tunggal, yang membedakannya adalah pada pautan tunggal menggunakan jarak minimum, sedangkan pada pautan lengkap

menggunakan jarak maximum, pautan rata- rata didasarkan atas observasi dengan jarak paling mendekati jarak rata – rata, metode ward didasarkan atas jumlah kuadrat antara dua gerombol untuk seluruh variable , dan metode centorid didasarkan atas jarak antar centroid gerombol tersebut.

9.2.3. Metode tak berhirarki

Salah satu metode dalam metode penggerombolan tak berhirarki yaitu metode *k-means*. Algoritma dari metode ini sebagai berikut pertama tentukan besarnya *k* (yaitu banyaknya gerombol, dan tentukan juga centroid di tiap gerombol), kedua hitung jarak antara setiap objek dengan setiap centroid, ketiga hitung kembali rataan (centroid) untuk gerombol yang baru terbentuk dan keempat ulangi langkah 2 sampai tidak ada lagi pemindahan objek antar gerombol.

Dua masalah utama yang harus diketahui dalam penggerombolan non hierarki adalah jumlah gerombol dan pemilihan pusat gerombol (*centroid*). Lebih lanjut, hasil penggerombolan mungkin tergantung pada pusat (centers) dipilih. Banyak program pengerombolan non-hierarki, memilih k objek (kasus) yang pertama sebagai pusat gerombol (centroid). Sehingga hasil penggerombolan mungkin tergantung pada urutan observasi dalam data. Bagaimanapun juga, penggerombolan non-hierarki lebih cepat daripada metode hirarki dan lebih menguntungkan jika jumlah objek/kasus atau observasi besar sekali(sampel besar).

9.2.3.1. *Metode K-Rataan*

MacQueen menyarankan penggunaan K-rataan untuk menjelaskan algoritma dalam penentuan suatu objek ke dalam gerombol tertentu berdasarkan rataan terdekat. Dalam bentuk yang paling sederhana, proses ini terdiri dari tiga tahap:

1. Bagi objek-objek tersebut ke dalam K gerombol awal.
2. Masukkan tiap objek ke suatu gerombol berdasarkan rataan terdekat. Jarak biasanya ditentukan dengan menggunakan Euclidean. Hitung kembali rataan untuk gerombol yang mendapat objek dan yang kehilangan objek.
3. Ulangi langkah 2 sampai tidak ada lagi pemindahan objek antar gerombol.

Dalam membagi objek ke dalam K kelompok permulaan (pada langkah 1), sebelumnya dapat ditentukan rataan untuk K inisial, baru kemudian dilanjutkan dengan langkah berikutnya.

Penentuan terakhir suatu objek ke suatu gerombol tertentu tidak tergantung dari K inisial yang pertama kali ditentukan. Pengalaman menunjukkan bahwa perubahan terbesar hanya terjadi pada realokasi yang pertama saja.

9.2.3.2. *Ilustrasi Metode penggerombolan hierarkhi*

Misalkan ada dua variabel X_1 dan X_2 yang tiap objeknya diberi inisial A, B, C dan D. Akan dilakukan pengelompokan menggunakan metode K-rataan. Datanya sebagai berikut

Objek	Pengamatan	
	x_1	x_2
A	5	3
B	-1	1
C	1	-2
D	-3	-2

Objek-objek diatas dibagi menjadi 2 gerombol (K=2) sedemikan sehingga antar objek dalam satu gerombol mempunyai hubungan yang lebih

dekat dari pada objek antar gerombol. Dengan menggunakan metode K=2 rataan, pembagian objek menjadi 2 gerombol , seperti (AB) dan (CD) dan kemudian menghitung koordinat ($\bar{x}_1$, $\bar{x}_2$) sebagai rataan dari gerombol. Dari langkah pertama dihasilkan:

Cluster	Koordinat Rataan	
	$\bar{x}_1$	$\bar{x}_2$
(AB)	$\frac{5+(-1)}{2}=2$	$\frac{3+(1)}{2}=2$
(CD)	$\frac{1+(-3)}{2}=-1$	$\frac{-2+(-2)}{2}=-2$

Pada langkah ke 2, dengan menggunakan perhitungan Euclidean dihitung jarak setiap objek dari pusat gerombol (rataannya) dan memasukkannya ke dalam gerombol terdekat. Pada setiap pemindahan objek pada suatu gerombol tertentu, K rataannya harus dihitung kembali sebelum dilanjutkan ke proses berikutnya. Perhitungan jaraknya sebagai berikut:

$d^2(A,(AB)) = (5-2)^2+(3-2)^2=10$

$d^2(A,(CD)) = (5+1)^2+(3+2)^2=61$

Karena jarak A lebih dekat dengan (AB) daripada dengan (CD), maka A tidak perlu pindah dari gerombol (AB). Dilanjutkan dengan :

$d^2(B,(AB)) = (-1+1)^2+(1+2)^2 =10$

$d^2(B,(CD)) = (-1+1)^2+(1+2)^2 = 9$

Karena jarak B lebih dekat ke (CD) dari pada ke (AB) , maka perlu ada pemindahan (B) ke gerombol (CD) sehingga sekarang terbentuk suatu gerombol baru yaitu (BCD) dengan koordinat baru.

Cluster	Koordinat Rataan	
	$\bar{x}_1$	$\bar{x}_2$
(A)	5	3
(BCD)	-1	-0.5

Kemudian periksa lagi setiap objek tersebut. Perhitungannya sebagai berikut:

Kuadrat Jarak ke Rataan Grup

Gerombol	Objek			
	A	B	C	D
(A)	0	40	41	89
(BCD)	48.25	2.25	6.25	6.25

Dapat dilihat sekarang bahwa setiap objek sudah menunjuk ke suatu gerombol dengan rataan terdekat sehingga proses berhenti sampai disini. Gerombol akhir yang terbentuk adalah A dan (BCD).

Untuk memeriksa kestabilan dari pembentukan gerombol disarankan untuk mengulang kembali langkah-langkah tersebut diatas dengan inisial gerombol awal yang berbeda.

9.3. Aplikasi SAS

Prosedur SAS yang digunakan untuk analisis gerombol adalah PROC CLUSTER.

Bentuk umum dari PROC CLUSTER adalah

```
PROC CLUSTER <options>;
        ID variables;
Run;
```

Berikut akan diberikan contoh penggunaan analisis gerombol berhirarki yang diterapkan pada penggerombolan mamalia berdasarkan struktur giginya. Mamalia yang terlibat adalah : BROWN BAT, MOLE, SILVER HAIR BAT, PIGMY BAT, HOUSE BAT, REDB AT, PIKA, RABBIT, BEAVER, GROUNDHOG, GRAYS QUIRREL, HOUSE MOUSE, PORCUPINE, WOLF, BEAR, RACCOON, MARTEN, WEASEL, WOLVERINE, BADGER, RIVER OTTER, SEA OTTER, JAGUAR, COUGAR, FURSEAL, SEALION, GREY SEAL, ELEPHANT SEAL, REINDEER, ELK, DEER, MOOSE

Berikut adalah data yang digunakan dalam analisis gerombol. Ada 8 peubah yang berkenaan dengan gigi mamalia, yaitu V1 hingga V8.

```
title 'HIERARCHICAL CLUSTER ANALYSIS OF MAMMALS'' TEETH
DATA';
title2 'Evaluating the Effects of Ties';
data teeth;
   input mammal $ 1-16@21 (v1-v8) (1.);
   label v1='Top incisors'
         v2='Bottom incisors'
         v3='Top canines'
         v4='Bottom canines'
         v5='Top premolars'
         v6='Bottom premolars'
         v7='Top molars'
         v8='Bottom molars';
   datalines;
BROWN BAT           23113333
MOLE                32103333
SILVER HAIR BAT     23112333
PIGMY BAT           23112233
HOUSE BAT           23111233
RED BAT             13112233
PIKA                21002233
RABBIT              21003233
BEAVER              11002133
GROUNDHOG           11002133
GRAY SQUIRREL       11001133
HOUSE MOUSE         11000033
PORCUPINE           11001133
WOLF                33114423
BEAR                33114423
RACCOON             33114432
MARTEN              33114412
WEASEL              33113312
```

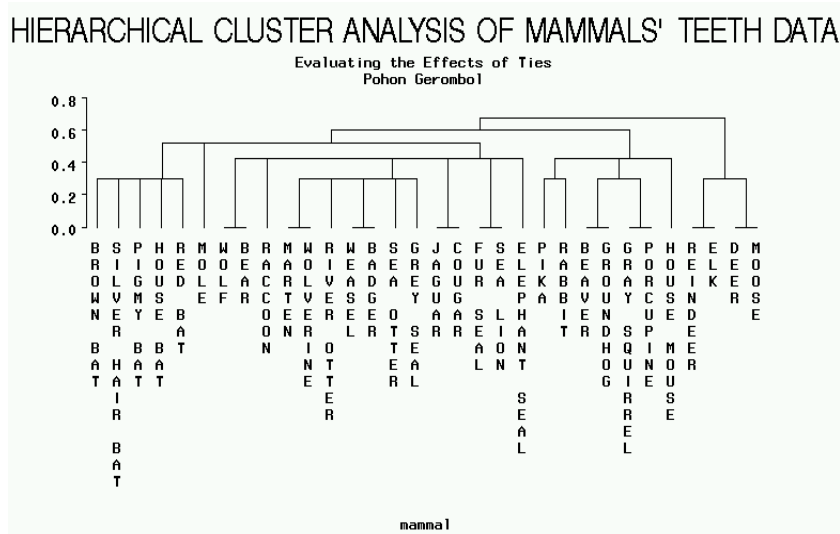
```
WOLVERINE           33114412
BADGER              33113312
RIVER OTTER         33114312
SEA OTTER           32113312
JAGUAR              33113211
COUGAR              33113211
FUR SEAL            32114411
SEA LION            32114411
GREY SEAL           32113322
ELEPHANT SEAL       21114411
REINDEER            04103333
ELK                 04103333
DEER                04003333
MOOSE               04003333
;

proc cluster data=teeth CCC method=single outtree= tree;
   var v1-v8;
   id mammal;
run;

proc tree;
   id mammal;
   title3 'Pohon Gerombol';
run;
```

Keluaran dari perintah di atas adalah :

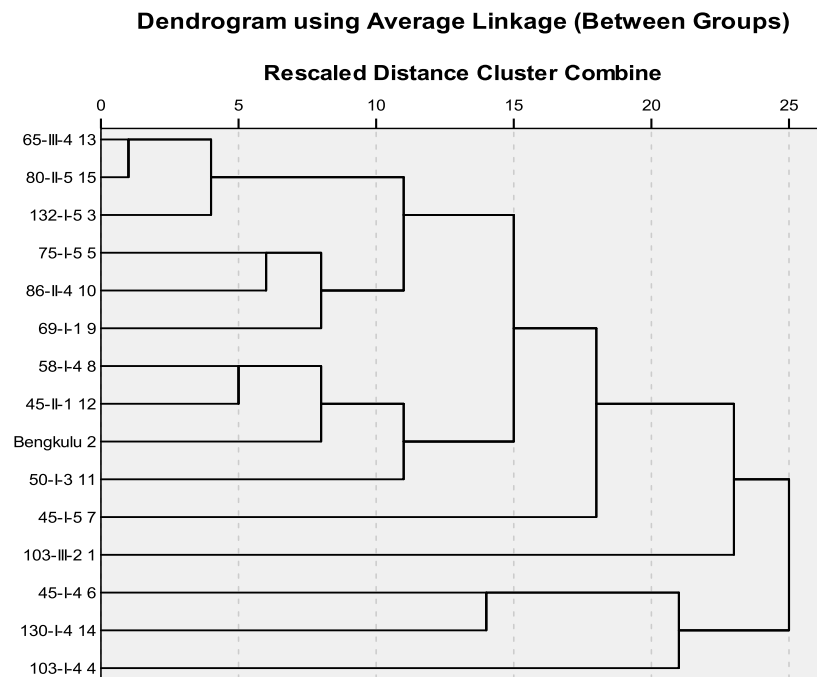
Cluster History

NCL	---------Clusters Joined----------		FREQ	SPRSQ	RSQ	ERSQ	CCC	Dist	Norm T Min i e
31	BEAVER	GROUNDHOG	2	0.0000	1.00	.	.	0	T
30	GRAY SQUIRREL	PORCUPINE	2	0.0000	1.00	.	.	0	T
29	WOLF	BEAR	2	0.0000	1.00	.	.	0	T
28	MARTEN	WOLVERINE	2	0.0000	1.00	.	.	0	T
27	WEASEL	BADGER	2	0.0000	1.00	.	.	0	T
26	JAGUAR	COUGAR	2	0.0000	1.00	.	.	0	T
25	FUR SEAL	SEA LION	2	0.0000	1.00	.	.	0	T
24	REINDEER	ELK	2	0.0000	1.00	.	.	0	T
23	DEER	MOOSE	2	0.0000	1.00	.	.	0	
22	BROWN BAT	SILVER HAIR BAT	2	0.0025	.998	.	.	0.301	T
21	CL22	PIGMY BAT	3	0.0042	.993	.	.	0.301	T
20	CL21	HOUSE BAT	4	0.0083	.985	.	.	0.301	T
19	CL20	RED BAT	5	0.0050	.980	.	.	0.301	T
18	PIKA	RABBIT	2	0.0025	.978	.	.	0.301	T
17	CL31	CL30	4	0.0050	.973	.	.	0.301	T
16	CL28	RIVER OTTER	3	0.0033	.969	.	.	0.301	T
15	CL16	CL27	5	0.0087	.961	.	.	0.301	T
14	CL15	SEA OTTER	6	0.0063	.954	.	.	0.301	T
13	CL14	GREY SEAL	7	0.0088	.945	.	.	0.301	T
12	CL24	CL23	4	0.0050	.940	.	.	0.301	
11	CL18	CL17	6	0.0200	.920	.	.	0.4256	T
10	CL11	HOUSE MOUSE	7	0.0225	.898	.	.	0.4256	T
9	CL29	RACCOON	3	0.0067	.891	.	.	0.4256	T
8	CL9	CL13	10	0.0292	.862	.	.	0.4256	T
7	CL8	CL26	12	0.0362	.826	.	.	0.4256	T
6	CL7	CL25	14	0.0230	.803	.779	1.25	0.4256	T
5	CL6	ELEPHANT SEAL	15	0.0252	.778	.741	1.50	0.4256	

4	CL19	MOLE	6	0.0200	.758	.691	2.13	0.5213	T
3	CL4	CL5	21	0.1722	.585	.614	-.73	0.5213	
2	CL3	CL10	28	0.4033	.182	.465	-4.2	0.6019	
1	CL2	CL12	32	0.1821	.000	.000	0.00	0.673	

Terlihat bahwa nilai CCC tertinggi ketika banyaknya gerombol adalah 6, namun nilai CCC ini pada saat banyaknya gerombol 5 tidak berbeda jauh. Sekarang kita perhatikan bagaimana bentuk dari pohon gerombolnya.

Jika kita amati, banyaknya gerombol 5 cukup memuaskan. Gerombol yang pertama merupakan mamalia dari keluarta BAT (kelelawar), gerombol berikutnya adalah kelompok mamalia pemakan daging, kemudian ada gerombol mamalia pengerat, gerombol mamalia pemakan rumbuhan dan terakhir MOLE. Nampak jelas bahwa penggerombolan ini cukup masuk akal.

9.4. Latihan

1. Didapatkan hasil dendogram sebagai berikut:

   a. Berdasarkan hasil diatas, ada berapa gerombol yang terbentuk?
   b. Sebutkan anggota dari masing-masing gerombol!

2. Diketahui matriks jarak

$$D = \begin{matrix} A \\ B \\ C \\ D \\ E \end{matrix} \begin{bmatrix} 0 & 9 & 3 & 6 & 11 \\ 9 & 0 & 7 & 5 & 10 \\ 3 & 7 & 0 & 9 & 2 \\ 6 & 5 & 9 & 0 & 8 \\ 11 & 10 & 2 & 8 & 0 \end{bmatrix}$$

Kelompokkan dan gambarkan dalam bentuk dendogram serta bandingkan hasilnya dari 5 pengamatan A,B,C,D, dan E menggunakan metode penggerombolan berikut:

a. Single Linkage
b. Complete Linkage
c. Average Linkage

3. Diketahui korelasi diantara 5 toko penyedia bahan makanan pokok di kota Bogor adalah sebagai berikut

	Sumber Makmur	Sri Rejeki	Sentosa	Victory Jaya	Sriwangi
Sumber Makmur	1				
Sri Rejeki	0.58	1			
Sentosa	0.51	0.6	1		
Victory Jaya	0.39	0.39	0.44	1	
Sriwangi	0.46	0.32	0.43	0.52	1

a. Gerombolkan toko-toko tersebut dengan metode single linkage dan compplete linkage!
b. Gambarkan dendogram dan bandingkan hasilnya.

4. Didapatkan pengukuran peubah X_1 dan X_2 untuk objek A,B,C dan D sebagai berikut

	Observsasi	
	X_1	X_2
A	5	4
B	1	-2
C	-1	1
D	3	1

a. Gunakam K-means clustering untuk membagi objek menjadi 2 kelompok K=2 dengan grup inisial awal adalah (AB) dan (CD)
b. Ulangi sepertipada soal a tetapi grup inisial awal adalah (AC) dan (BD)
c. Bandingkan kedua hasilnya.

10

10.Analisis Diskriminan (*Discriminant Analysis*)

10.1. Pendahuluan

Analisis diskriminan adalah salah satu teknik statistik yang bisa digunakan pada hubungan dependensi (hubungan antarvariabel dimana sudah bisa dibedakan mana variabel respon dan mana variabel penjelas). Lebih spesifik lagi, analisis diskriminan digunakan pada kasus dimana variabel respon berupa data kualitatif dan variabel penjelas berupa data kuantitatif. Analisis diskriminan bertujuan untuk mengklasifikasikan suatu individu atau observasi ke dalam kelompok yang saling bebas (mutually exclusive/disjoint) dan menyeluruh (exhaustive) berdasarkan sejumlah variabel penjelas.

Ada dua asumsi utama yang harus dipenuhi pada analisis diskriminan ini, yaitu:

2. Sejumlah p variabel penjelas harus berdistribusi normal.
3. Matriks varians-covarians variabel penjelas berukuran pxp pada kedua kelompok harus sama.

Jika dianalogikan dengan regresi linier, maka analisis diskriminan merupakan kebalikannya. Pada regresi linier, variabel respon yang harus mengikuti distribusi normal dan homoskedastis, sedangkan variabel

penjelas diasumsikan fixed, artinya variabel penjelas tidak disyaratkan mengikuti sebaran tertentu. Untuk analisis diskriminan, variabel penjelasnya seperti sudah disebutkan di atas harus mengikuti distribusi normal dan homoskedastis, sedangkan variabel responnya fixed.

10.2. Model Analisis Diskriminan

Model dasar analisis diskriminan mirip regresi berganda. Perbedaannya adalah kalau variable dependen regresi berganda dilambangkan dengan Y, maka dalam analisis diskriminan dilambangkan dengan D.

Model dasar analisis diskriminan adalah sebuah persamaan yang menunjukkan suatu kombinasi linear dari berbagai variable independent, yaitu :

$$D = b_0 + b_1X_1 + b_2X_2 + b_3X_3 + \cdots b_kX_k$$

Dimana :

D = skor diskriminan

b = koefisien diskriminan atau bobot

X = predictor atau variable independen

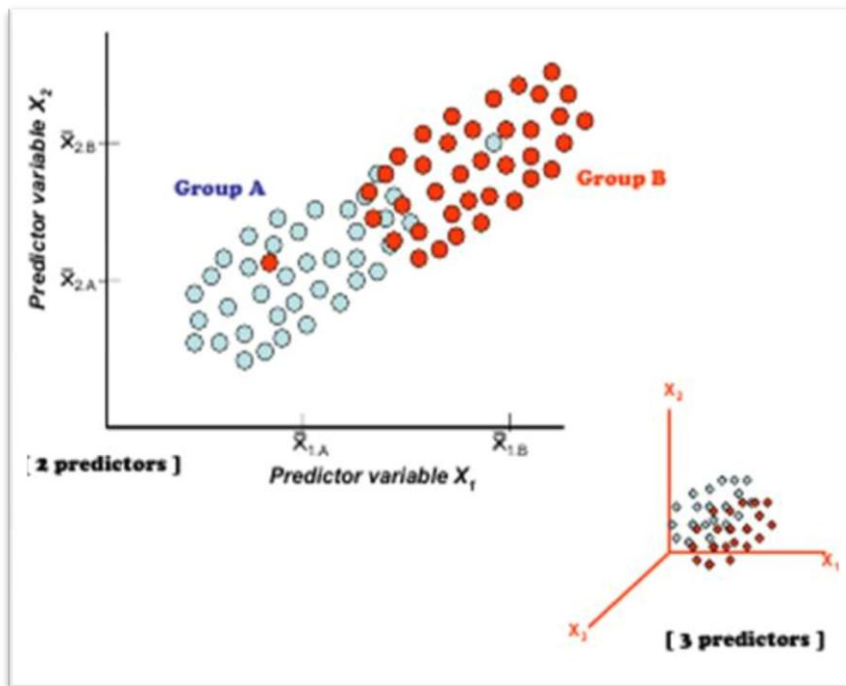
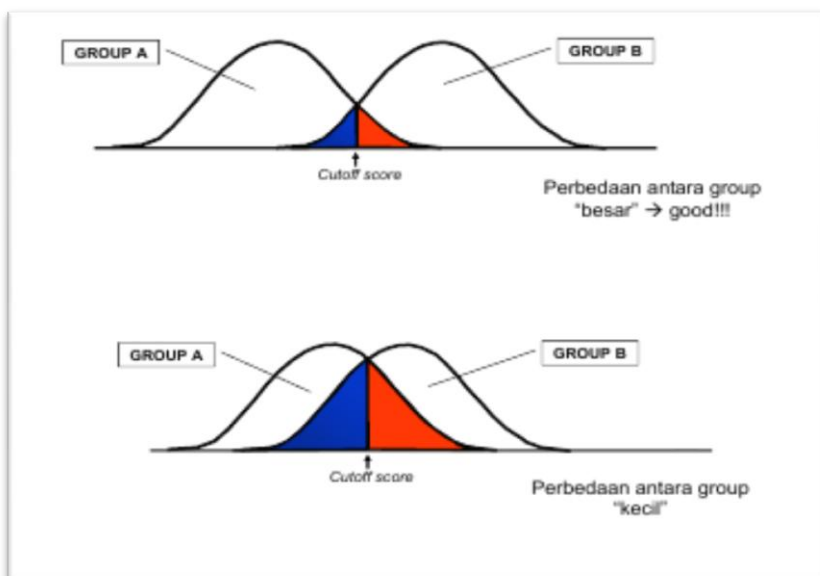
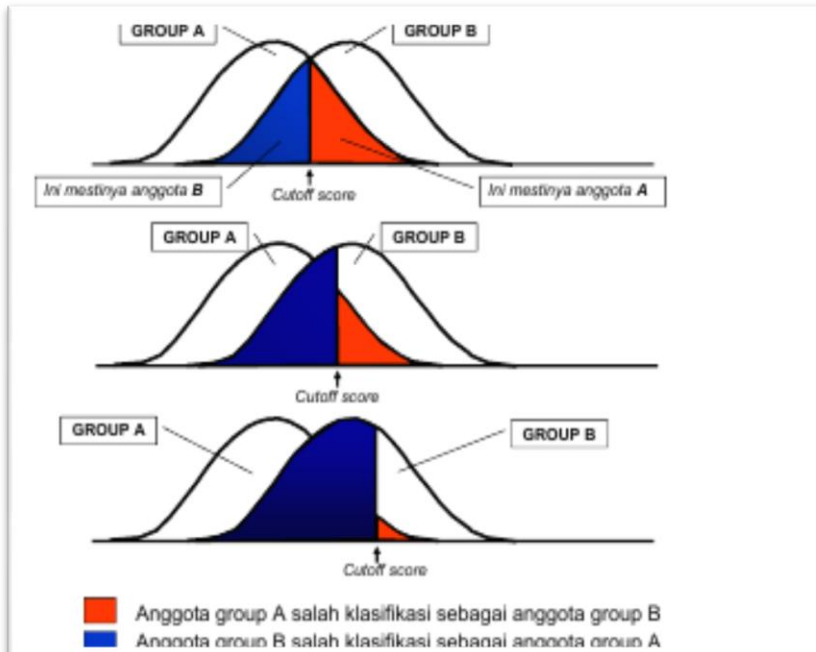
Yang diestimasi adalah koefisien ‘b’, sehingga nilai ‘D’ setiap grup sedapat mungkin berbeda. Ini terjadi pada saat rasio jumlah kuadrat antargrup (*between-group sum of square*) terhadap jumlah kuadrat dalam grup (*within-group sum of square*) untuk skor diskriminan mencapai maksimum. Berdasarkan nilai D itulah keanggotaan seseorang diprediksi.

Merangkum banyak peubah menjadi sebuah indeks atau menjadi sebuah criteria yang terandal akan sangat membantu proses klasifikasi. Namun perlu diingat bahwa kaidah keputusan pengklasifikasian berdasarkan indeks ataupun criteria apapun tidak selalu bisa diharapkan memiliki ketepatan yang sempurna. Artinya, dengan penyusunan indeks atau kriteria ini tetap selalu ada peluang kesalahan klasifikasi. Sewajarnya peluang kesalahan ini yang akan dikendalikan sehingga bisa dibuat sekecil mungkin.

Fungsi Diskriminan memiliki ukuran yang menggambarkan tingkat ketepatan :

Gambar Deskripsi umum analisis diskriminan

GROUP A
GROUP B
Ini mestinya anggota B
Cutoff score
Ini mestinya anggota A
GROUP A
GROUP B
Cutoff score
GROUP A
GROUP B
Cutoff score
Anggota group A salah klasifikasi sebagai anggota group B
Anggota group B salah klasifikasi sebagai anggota group A

GROUP A
GROUP B
Cutoff score
Perbedaan antara group "besar" → good!!!
GROUP A
GROUP B
Cutoff score
Perbedaan antara group "kecil"

Ketika kita diminta untuk memprediksi individu mana termasuk ke dalam kelompok tertentu atau ketika kita diminta untuk mengidentifikasi sifat-sifat umum anggota suatu kelompok, maka kita berhadapan dengan persoalan pengelompokan dan penentuan sifat-sifat khas suatu kelompok. Misalnya, suatu wilayah dikatakan perkotaan, paling tidak, kalau (1) penduduknya banyak, (2) mempunyai banyak fasilitas, dan (3) kegiatan ekonomi penduduknya beragam. Sifat khas (1) dan (2) pada umumnya dapat dilihat secara kasat mata akan tetapi tidak pada sifat khas (3). Contoh lain, bagaimana kita dapat menentukan apakah seseorang miskin atau tidak miskin. Sifat umum dari orang miskin yang dapat dikenali barangkali adalah (1) kualitas rumah tinggalnya yang rendah, (2) porsi pengeluaran untuk makanan sangat besar (lebih dari 80 persen), dan (3) tingkat pendidikannya rendah.

Dalam teknik statistik, persoalan di atas biasanya diatasi dengan menggunakan analisis diskriminan.[11] Dua hal, yaitu pengelompokan dan identifikasi sifat khas suatu kelompok, dapat dilakukan sekaligus dengan analisis tersebut, di mana kelompok dikenal sebagai *group* dan sifat khas dikenal sebagai variabel pembeda (*discriminating variables*). Antara kelompok dan variabel pembeda tersebut kemudian dibuat suatu hubungan fungsional yang disebut dengan fungsi diskriminan.

---

[11] Dalam literatur statistik lebih banyak disebut sebagai analisis fungsi diskriminan.

10.3. Fungsi Diskriminan

Analisis ini didasarkan atas fungsi diskriminan yang mempunyai bentuk umum:

$$\hat{Y}_{i.} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p + \varepsilon_i$$

di mana Y merupakan variabel boneka (*dummy variables*) yang menunjukkan kelompok dan X_i adalah variabel pembeda. Pada dasarnya fungsi diskriminan merupakan fungsi regresi ganda dengan variabel dependen merupakan variabel boneka yang mengambil nilai 1, 2, ..., k yang sesuai dengan pengelompokan awal setiap individu dan banyaknya kelompok (=k).

Persyaratan awal yang harus dipenuhi sebelum melakukan analisis dengan fungsi diskriminan adalah:

1. Setiap individu harus dikelompokkan hanya ke dalam satu dan hanya satu kelompok
2. Varians dalam setiap kelompok adalah sama (*equal variances*)
3. X berdistribusi normal ganda (*multi variates normal distribution*)
4. Banyaknya kelompok harus memenuhi 2 k<p, dengan p adalah banyaknya variabel pembeda.[13]

Analisis diskriminan dapat dilakukan bila terdapat perbedaan yang nyata antar kelompok, sehingga pada tahap awal yang harus dilakukan adalah

---

[13] Pada umumnya sangat sulit sekali untuk dapat memenuhi persyaratan (2) dan (3), yang dalam praktek tidak pernah diuji; halmana akan membuat akurasi dari analisis dengan fungsi diskriminan akan berkurang. Namun demikian, fungsi diskriminan selalu menghasilkan estimasi yang kokoh (*robust estimates*) terutama yang berkaitan dengan prediksi pengelompokan.

uji hipotesis nol bahwa tidak ada perbedaan kelompok di antara individu yang dirumuskan dengan:

Ho: $\mu_1 = \mu_2 = \ldots\ldots\ldots = \mu_k$
H1: $\mu_1 \neq \mu_j$ si,j=1,2,...,k (sedikitnya ada 2 kelompok yang berbeda).

di mana

$\mu_j = [\mu_{1j}, \mu_{2j}, \ldots, \mu_{pj}]'$ adalah vektor rata-rata hitung peubah pembeda pada populasi ke-j dan jumlah individu yang dianalisis $N = \Sigma N_i$, maka probabilita pengelompokan awal (*prior probability*) adalah N_i/N.

Uji hipotesis tersebut dilakukan dengan menggunakan statistik Khi Kuadrat yang diperoleh dari relasi dengan Wilk lambda (λ)[14] yang menguji perbedaan antara dua kelompok. Jika H_0 benar maka $V_j = \{N - 1 - ½ (p+k)\}$ *ln* $(1/\lambda_j)$ mengikuti distribusi Khi kuadrat dengan derajat bebas p(k-1). Pada *output* SPSS hal ini dapat dilihat dari Tabel B berikut.

Tabel B:

DISCRIMINANTFUNCTION	EIGEN VALUES	RELATIVE PERCENTAGE	CANONICAL CORRELATION	FUNCTION DERIVED	WILK'S LAMBDA	CHI-SQUARE	DF	SIG.
				0	0.6329	85.538	12	0.00
1	0.31781	61.52	0.491	1	0.8341	33.931	6	0.000
2	0.19802	38.33	0.407	2	0.9992	0.145	2	0.930
3	0.00078	0.15	0.028					

[14] Dimana $\lambda_i = 1/(1+e_j)$, dan e_j adalah akar karakteristik (*eigen value*) ke–j dari BW^{-1}. B adalah variasi/jumlah kuadrat antarkelompok (*between groups sum of squares*) dan W adalah variasi/jumlah kuadrat dalam kelompok (*within groups sum of squares*).

Sebelum fungsi diskriminan dibentuk -- ***FUNCTION DERIVED = 0*** -- tampak bahwa antarkelompok berbeda secara nyata dengan tingkat signifikansi sebesar 0,000. Dengan perkataan lain H_0 ditolak.

Karena tujuan utama adalah pengelompokan, maka pembentukan fungsi diskriminan dicapai dengan menentukan β_j, j = 1,2,3,…,p -- dengan proses matematika -- sedemikian rupa sehingga perbedaan antarkelompok sangat jelas. Banyaknya fungsi diskriminan yang dibentuk tergantung dari banyaknya kelompok yang diidentifikasi, yang secara umum adalah sebanyak k–1. Fungsi diskri-minan yang dibentuk merupakan kombinasi linier dari peubah pembeda yang dapat membedakan kelompok secara maksimum.

10.4. Signifikansi Fungsi Diskriminan

Setelah dapat ditunjukkan bahwa Ho: $\mu_1 = \mu^2 = \ldots\ldots\ldots = \mu_k$ ditolak, maka sebanyak k-1 fungsi diskriminan dapat dibentuk. Untuk mengetahui apakah k–1 fungsi diskriminan tersebut signifikan untuk membedakan k kelompok sehingga layak untuk digunakan untuk analisis selanjutnya, suatu uji dilakukan di mana statistik Bartlett $V_j = \{N - 1 - ½ \ (p+k)\}\ \mathit{ln}\ (1+e_j)$ dengan $N = \Sigma N_l$ akan mengikuti distribusi Khi Kuadrat dengan derajat bebas p(k-1) jika H_0 benar. Contoh hasil uji dapat dilihat pada Tabel B di mana dari 3 fungsi diskriminan yang terbentuk untuk membedakan 4 kelompok hanya 2 fungsi diskriminan saja yang siginifikan untuk membedakan 4 kelompok.

Untuk analisis diskriminan untuk membedakan dua kelompok uji hipotesis untuk mengetahui signifikansi fungsi diskriminan, statistik V Bartlett menjadi

lebih sederhana, yaitu: V={N 1 ½(p+2)}*ln*(1+e) mengikuti distribusi Khi-Kuadrat dengan derajat bebas p, jika H_0 benar.

Menurut Johnson & Wichern (1982), tujuan fungsi pembeda adalah untuk menggambarkan ciri-ciri suatu pengamatan dari bermacam–macam populasi yang diketahui, baik secara grafis ataupun secara aljabar dengan membentuk fungsi pembeda.

Fungsi pembeda layak dibentuk bila terdapat perbedaan nilai rata–rata diantara kelompok-kelompok yang ada, oleh karena itu sebelum fungsi pembeda dibentuk perlu dilakukan pengujian terhadap perbedaan vektor nilai rata–rata dari kelompok-kelompok tersebut.

Seperti telah dijelaskan sebelumnya, dalam pengujian vektor nilai rata–rata antarkelompok, asumsi yang harus dipenuhi adalah:

1. Peubah-peubah yang diamati menyebar secara normal ganda (*multivariate normality*).
2. Semua kelompok populasi mempunyai matrik kovarians Σ yang sama (*equality of covariance matrices*).

Berikut ini, diuraikan bagaimana melakukan pengujian kedua asumsi di atas secara manual dengan memanfaatkan rumus.

10.4.1. Uji Kenormalan Multivariat

Menurut Karson (1982), untuk menguji kenormalan multivariat digunakan prosedur yang dikembangkan oleh Mardia (1970) dengan cara menghitung dua macam ukuran statistik yaitu ukuran *skewness* ($b_{1,p}$) dan *kurtosis* ($b_{2,p}$) yaitu:

$$b_{1,p} = \frac{1}{n^2}\sum\sum[(X_i - \bar{X})'S^{-1}(X_i - \bar{X})]^3$$

$$b_{2,p} = \frac{1}{n^2}\sum[(X_i - \bar{X})'S^{-1}(X_i - \bar{X})]^2$$

Hipotesis yang digunakan adalah:

H_o: multivariat mengikuti sebaran normal
H_1: multivariat tidak mengikuti sebaran normal

Bila ($b_{1,p}$ n/6 $\leq \chi^2$ p(p+1) (p+2)/6, dan
[$b_{2,p}$ - p (p+2)] / √ (8p (p+2)/n) $\leq Z_\alpha$ (tabel normal), maka Ho diterima, berarti multivariat mengikuti sebaran normal.

Menurut Johnson (1982), untuk menguji kenormalan ganda adalah dengan mencari nilai jarak kuadrat untuk setiap pengamatan yaitu

$$d_i^2 = (X_i - \bar{X})'S^{-1}(X_i - \bar{X})$$

di mana X_i adalah pengamatan yang ke –i dan S^{-1} adalah kebalikan (*inverse*) matrik kovarians S.

Kemudian d_i^2 diurutkan dari kecil ke besar, selanjutnya dibuat *plot* d_i^2 dengan nilai Chi-Kuadrat χ^2(i–1/2)/n , di mana i = urutan = 1, 2, ……n, dan p = banyaknya peubah. Bila hasil *plot* dapat didekati dengan garis lurus, maka dapat disimpulkan bahwa data menyebar secara normal ganda.

Menurut Norusis (1986), berdasarkan teori Wahl & Kronmal (1977), seringkali kenormalan ganda sulit diperoleh terutama bila sampel yang diambil relatif kecil. Bila hal ini terjadi, uji vektor rata–rata tetap bisa dilakukan selama asumsi kedua (kesamaan kovarians) dipenuhi.

10.4.2. Uji Kesamaan Matriks Kovarians

Untuk menguji kesamaan matriks kovarians (Σ) antar kelompok, digunakan hipotesa:

Ho: $\Sigma_1 = \Sigma_2 = \ldots\ldots = \Sigma_k = \Sigma$
H1: sedikitnya ada 2 kelompok yang berbeda

Statistik uji yang digunakan adalah statistik Box's, yaitu:
$-2 \ln \lambda^* = (n-k) \ln |W/(n-k)| - \Sigma (ni-1) \ln |S_j|$

di mana:

$$\lambda^* = \frac{\prod |S_j|^{(n_j-1)/2}}{|W/(n-k)|^{(n-k)/2}}$$

k = banyaknya kelompok

W/(n-k) = matriks kovarians dalam kelompok gabungan

Sj = matriks kovarians kelompok ke-j

Bila hipotesa nol benar, maka:

$(-2 \ln \lambda^*)/b$ akan mengikuti sebaran F dengan derajat besar v_1 dan v_2 pada taraf nyata α, di mana:

$v_1 = (1/2)(k-1)\,p\,(p+1)$

$v_2 = (v_1+2)/(a_2 - a_1^2)$

$b = v_1/(1-a_1 - v_1/v_2)$

$$a_1 = \frac{2p^2+3p-1}{6(k-1)(p+1)}\left[\sum \frac{1}{(n_j-1)} - \frac{1}{(n-k)}\right]$$

$$a_2 = \frac{(p-1)(p+2)}{6(k-1)}\left[\sum \frac{1}{(n_j-1)^2} - \frac{1}{(n-k)^2}\right]$$

p = jumlah peubah pembeda dalam fungsi pembeda (*discriminant*).

Karena itu, apabila (-2 ln λ^*)/b $\leq F_{v_1,v_2,\alpha}$ maka tidak ada alasan untuk menolak Ho dan dapat disimpulkan bahwa semua kelompok mempunyai matrik kovarians yang sama. Sebaliknya bila (-2 ln λ^*)/b > $F_{v_1,v_2,\alpha}$ maka Ho ditolak.

10.4.3. Uji Vektor Nilai Rata–rata

Pengujian terhadap vektor nilai rata–rata antarkelompok dilakukan dengan hipotesa:

Ho : $\mu_1 = \mu^2 = \ldots\ldots\ldots = \mu_k$
H1 : sedikitnya ada 2 kelompok yang berbeda.

Statistik uji yang digunakan adalah statistik V-Bartlett, yang mengikuti sebaran Chi-Kuadrat dengan derajat bebas p(k-1), bila bipotesa nol benar.

Statistik V-Bartlett diperoleh melalui:

$V = - [(n-1) – (p+k)/2] \ln (\Lambda)$

di mana:

n = banyaknya pengamatan
p = banyaknya peubah dalam fungsi pembeda (*discriminant*)
k = banaknya kelompok

$$\Lambda = \frac{|W|}{|B + W|}$$

Dalam hal ini:

Λ = Wilk's Lamda

W = matrik jumlah kuadrat dan hasil kali data dalam kelompok
B = matrik jumlah kuadrat dan hasil kali data antar kelompok

X_{ij} = pengamatan ke-j pada kelompok ke-i

$\bar{X}_i$ = vektor rata–rata kelompok ke-i

n_i = jumlah pengamatan pada kelompok ke-i

$\bar{X}$ = vektor rata–rata total

Apabila $V \leq \chi 2\ p(k\text{-}1),(1\text{-}\alpha)$, maka tidak ada alasan untuk menolak Ho, ini berarti bahwa terdapat perbedaan vektor nilai rata–rata antarkelompok. Bila $V > \chi 2\ p(k\text{-}1),\ (1\text{-}\alpha)$, maka Ho ditolak.

Bila dari hasil pengujian di atas ternyata ada perbedaan vektor nilai rataan, maka fungsi pembeda (*discriminant*) layak disusun untuk mengkaji hubungan antarkelompok serta berguna untuk mengelompokkan suatu obyek baru ke dalam salah satu kelompok tersebut.

10.5. Variabel Pembeda Utama

Untuk memperoleh fungsi diskriminan dengan SPSS digunakan **METHOD=DIRECT**. Dengan method ini, SPSS akan memberikan fungsi diskriminan yang merupakan kombinasi linier dari semua variabel pembeda sebagaimana pada persamaan (4.1). Namun demikian, dengan uji parsial signifikansi variabel pembeda akan dapat diperoleh variabel-variabel mana yang sebenarnya tidak berpengaruh nyata dalam membedakan kelompok sehingga tidak perlu ada dalam fungsi diskriminan.

Penentuan variabel pembeda yang siginifikan dalam membedakan kelompok dapat dilakukan dengan cara bertatar (*stepwise*) yang prinsipnya sudah dibahas dalam Bab Analisis Regresi. Dalam SPSS, untuk memperoleh fungsi diskriminan yang merupakan kombinasi linier dari variabel pembeda utama dilakukan dengan cara menggunakan

METHOD[15] = *Wilks' lambda, Unexplained variance, Mahalanobis distance, Smallest F Ratio, Rao's V*. Dalam prosedur bertatar analisis diskriminan, setiap peubah pembeda yang tidak masuk dalam fungsi diskriminan berarti tidak menambah kemampuan membedakan kelompok (*discriminating power*) dari fungsi.

Variabel pembeda utama yang dihasilkan dari proses bertatar merupakan ciri umum dari semua kelompok yang diteliti. Misalnya, jika Y menyatakan apakah individu merupakan penduduk miskin atau bukan miskin yang diperoleh apakah pengeluaran per kapita sebulan dari rumah tangganya di perkotaan di atas Rp.95.000,- atau di atas Rp.75.000,- di pedesaan. Variabel pembeda (X) yang diselidiki misalnya kualitas rumah tinggal, pendidikan kepala rumah tangga, jumlah anggota rumah tangga, jam kerja seminggu dari kepala rumah tangga. Dalam hal ini variabel pembeda merupakan ciri dari penduduk miskin.

10.6. Evaluasi Fungsi Diskriminan

Hasil pengelompokan menurut fungsi diskriminan tidak selalu sama dengan pengelompokan awal. Besarnya kesalahan pengelompokan, dengan menganggap pengelompokan awal adalah benar, merupakan indikator tingkat akurasi dari fungsi diskriminan yang dihasilkan. Matriks 14.2 berikut menunjukkan evaluasi terhadap fungsi diskriminan.

---

[15] Untuk lebih mengetahui keunggulan dan kelemahan setiap motode bertatar, pembaca dianjurkan untuk membaca *SPSS Reference Guide*.

Matriks 14.2. Validasi hasil klasifikasi fungsi diskriminan

PENGELOMPOKAN AWAL	PENGELOMPOKAN MENURUT FUNGSI DISKRIMINAN		JUMLAH
	I	II	
I	N_{11}	N_{12}	$N_{1\bullet}$
II	N_{21}	N_{22}	$N_{2\bullet}$
JUMLAH	$N_{\bullet 1}$	$N_{\bullet 2}$	N

Dengan menggunakan Matriks 14.2 dapat dievaluasi tingkat akurasi fungsi diskriminan dengan memperhatikan:

(a). Persentase tepat pengelompokan = $(N_{11} + N_{12}) / N$

(b). Probabilita pengelompokan awal (*prior probability*) = $N_{1.} / N$

Fungsi diskriminan dikatakan cukup baik jika (a) $\geq$ 1,25 (b)

10.7. Aplikasi SAS

Ada tiga prosedur utama pada SAS untuk analisis diskriminan yakni:

a. PROC CANDISC untuk analisis diskriminan kanonik

b. PROC DISCRIM untuk analisis diskriminan linear

c. PROC STEPDISC untuk analisis diskriminan bertatar

PROC CANDISC selalu dikerjakan pertama kali. PROC DISCRIM yang kedua. Bentuk umum dari tahapan PROC CANDISC adalah

```
PROC CANDISC <options>
CLASS variable;
VAR variables;
RUN;
```

Berikut diberikan teladan penerapan analisis diskriminan untuk membuat fungsi pembeda antara 6 spesies ikan: **BREAM, ROACH, WHITEFISH, PARKKI, PERCH, PIKE, SMELT**. Peubah yang digunakan untuk menyusun fungsi diskriminan adalah BERAT TUBUH, PANJANG TUBUH, PANJANG KEPALA, PANJANG EKOR, TINGGI IKAN, LEBAR IKAN.

Tahapan penyiapan datanya adalah

```
proc format;
   value specfmt
      1='Bream'
      2='Roach'
      3='Whitefish'
      4='Parkki'
      5='Perch'
      6='Pike'
      7='Smelt';
run;

data fish (drop=HtPct WidthPct);
   title 'Fish Measurement Data';
   input Species Weight Length1 Length2 Length3 HtPct WidthPct @@;
   Height=HtPct*Length3/100;
   Width=WidthPct*Length3/100;
   format Species specfmt.;
   symbol = put(Species, specfmt.);
   datalines;
1  242.0 23.2 25.4 30.0 38.4 13.4 1  290.0 24.0 26.3 31.2 40.0 13.8
1  340.0 23.9 26.5 31.1 39.8 15.1 1  363.0 26.3 29.0 33.5 38.0 13.3
1  430.0 26.5 29.0 34.0 36.6 15.1 1  450.0 26.8 29.7 34.7 39.2 14.2
1  500.0 26.8 29.7 34.5 41.1 15.3 1  390.0 27.6 30.0 35.0 36.2 13.4
1  450.0 27.6 30.0 35.1 39.9 13.8 1  500.0 28.5 30.7 36.2 39.3 13.7
1  475.0 28.4 31.0 36.2 39.4 14.1 1  500.0 28.7 31.0 36.2 39.7 13.3
1  500.0 29.1 31.5 36.4 37.8 12.0 1      .  29.5 32.0 37.3 37.3 13.6
1  600.0 29.4 32.0 37.2 40.2 13.9 1  600.0 29.4 32.0 37.2 41.5 15.0
1  700.0 30.4 33.0 38.3 38.8 13.8 1  700.0 30.4 33.0 38.5 38.8 13.5
1  610.0 30.9 33.5 38.6 40.5 13.3 1  650.0 31.0 33.5 38.7 37.4 14.8
1  575.0 31.3 34.0 39.5 38.3 14.1 1  685.0 31.4 34.0 39.2 40.8 13.7
1  620.0 31.5 34.5 39.7 39.1 13.3 1  680.0 31.8 35.0 40.6 38.1 15.1
1  700.0 31.9 35.0 40.5 40.1 13.8 1  725.0 31.8 35.0 40.9 40.0 14.8
1  720.0 32.0 35.0 40.6 40.3 15.0 1  714.0 32.7 36.0 41.5 39.8 14.1
1  850.0 32.8 36.0 41.6 40.6 14.9 1 1000.0 33.5 37.0 42.6 44.5 15.5
1  920.0 35.0 38.5 44.1 40.9 14.3 1  955.0 35.0 38.5 44.0 41.1 14.3
1  925.0 36.2 39.5 45.3 41.4 14.9 1  975.0 37.4 41.0 45.9 40.6 14.7
1  950.0 38.0 41.0 46.5 37.9 13.7
2   40.0 12.9 14.1 16.2 25.6 14.0 2   69.0 16.5 18.2 20.3 26.1 13.9
2   78.0 17.5 18.8 21.2 26.3 13.7 2   87.0 18.2 19.8 22.2 25.3 14.3
2  120.0 18.6 20.0 22.2 28.0 16.1 2    0.0 19.0 20.5 22.8 28.4 14.7
2  110.0 19.1 20.8 23.1 26.7 14.7 2  120.0 19.4 21.0 23.7 25.8 13.9
2  150.0 20.4 22.0 24.7 23.5 15.2 2  145.0 20.5 22.0 24.3 27.3 14.6
2  160.0 20.5 22.5 25.3 27.8 15.1 2  140.0 21.0 22.5 25.0 26.2 13.3
```

```
2  160.0 21.1 22.5 25.0 25.6 15.2 2  169.0 22.0 24.0 27.2 27.7 14.1
2  161.0 22.0 23.4 26.7 25.9 13.6 2  200.0 22.1 23.5 26.8 27.6 15.4
2  180.0 23.6 25.2 27.9 25.4 14.0 2  290.0 24.0 26.0 29.2 30.4 15.4
2  272.0 25.0 27.0 30.6 28.0 15.6 2  390.0 29.5 31.7 35.0 27.1 15.3
3  270.0 23.6 26.0 28.7 29.2 14.8 3  270.0 24.1 26.5 29.3 27.8 14.5
3  306.0 25.6 28.0 30.8 28.5 15.2 3  540.0 28.5 31.0 34.0 31.6 19.3
3  800.0 33.7 36.4 39.6 29.7 16.6 3 1000.0 37.3 40.0 43.5 28.4 15.0
4   55.0 13.5 14.7 16.5 41.5 14.1 4   60.0 14.3 15.5 17.4 37.8 13.3
4   90.0 16.3 17.7 19.8 37.4 13.5 4  120.0 17.5 19.0 21.3 39.4 13.7
4  150.0 18.4 20.0 22.4 39.7 14.7 4  140.0 19.0 20.7 23.2 36.8 14.2
4  170.0 19.0 20.7 23.2 40.5 14.7 4  145.0 19.8 21.5 24.1 40.4 13.1
4  200.0 21.2 23.0 25.8 40.1 14.2 4  273.0 23.0 25.0 28.0 39.6 14.8
4  300.0 24.0 26.0 29.0 39.2 14.6
5    5.9  7.5  8.4  8.8 24.0 16.0 5   32.0 12.5 13.7 14.7 24.0 13.6
5   40.0 13.8 15.0 16.0 23.9 15.2 5   51.5 15.0 16.2 17.2 26.7 15.3
5   70.0 15.7 17.4 18.5 24.8 15.9 5  100.0 16.2 18.0 19.2 27.2 17.3
5   78.0 16.8 18.7 19.4 26.8 16.1 5   80.0 17.2 19.0 20.2 27.9 15.1
5   85.0 17.8 19.6 20.8 24.7 14.6 5   85.0 18.2 20.0 21.0 24.2 13.2
5  110.0 19.0 21.0 22.5 25.3 15.8 5  115.0 19.0 21.0 22.5 26.3 14.7
5  125.0 19.0 21.0 22.5 25.3 16.3 5  130.0 19.3 21.3 22.8 28.0 15.5
5  120.0 20.0 22.0 23.5 26.0 14.5 5  120.0 20.0 22.0 23.5 24.0 15.0
5  130.0 20.0 22.0 23.5 26.0 15.0 5  135.0 20.0 22.0 23.5 25.0 15.0
5  110.0 20.0 22.0 23.5 23.5 17.0 5  130.0 20.5 22.5 24.0 24.4 15.1
5  150.0 20.5 22.5 24.0 28.3 15.1 5  145.0 20.7 22.7 24.2 24.6 15.0
5  150.0 21.0 23.0 24.5 21.3 14.8 5  170.0 21.5 23.5 25.0 25.1 14.9
5  225.0 22.0 24.0 25.5 28.6 14.6 5  145.0 22.0 24.0 25.5 25.0 15.0
5  188.0 22.6 24.6 26.2 25.7 15.9 5  180.0 23.0 25.0 26.5 24.3 13.9
5  197.0 23.5 25.6 27.0 24.3 15.7 5  218.0 25.0 26.5 28.0 25.6 14.8
5  300.0 25.2 27.3 28.7 29.0 17.9 5  260.0 25.4 27.5 28.9 24.8 15.0
5  265.0 25.4 27.5 28.9 24.4 15.0 5  250.0 25.4 27.5 28.9 25.2 15.8
5  250.0 25.9 28.0 29.4 26.6 14.3 5  300.0 26.9 28.7 30.1 25.2 15.4
5  320.0 27.8 30.0 31.6 24.1 15.1 5  514.0 30.5 32.8 34.0 29.5 17.7
5  556.0 32.0 34.5 36.5 28.1 17.5 5  840.0 32.5 35.0 37.3 30.8 20.9
5  685.0 34.0 36.5 39.0 27.9 17.6 5  700.0 34.0 36.0 38.3 27.7 17.6
5  700.0 34.5 37.0 39.4 27.5 15.9 5  690.0 34.6 37.0 39.3 26.9 16.2
5  900.0 36.5 39.0 41.4 26.9 18.1 5  650.0 36.5 39.0 41.4 26.9 14.5
5  820.0 36.6 39.0 41.3 30.1 17.8 5  850.0 36.9 40.0 42.3 28.2 16.8
5  900.0 37.0 40.0 42.5 27.6 17.0 5 1015.0 37.0 40.0 42.4 29.2 17.6
5  820.0 37.1 40.0 42.5 26.2 15.6 5 1100.0 39.0 42.0 44.6 28.7 15.4
5 1000.0 39.8 43.0 45.2 26.4 16.1 5 1100.0 40.1 43.0 45.5 27.5 16.3
5 1000.0 40.2 43.5 46.0 27.4 17.7 5 1000.0 41.1 44.0 46.6 26.8 16.3
6  200.0 30.0 32.3 34.8 16.0  9.7 6  300.0 31.7 34.0 37.8 15.1 11.0
6  300.0 32.7 35.0 38.8 15.3 11.3 6  300.0 34.8 37.3 39.8 15.8 10.1
6  430.0 35.5 38.0 40.5 18.0 11.3 6  345.0 36.0 38.5 41.0 15.6  9.7
6  456.0 40.0 42.5 45.5 16.0  9.5 6  510.0 40.0 42.5 45.5 15.0  9.8
6  540.0 40.1 43.0 45.8 17.0 11.2 6  500.0 42.0 45.0 48.0 14.5 10.2
6  567.0 43.2 46.0 48.7 16.0 10.0 6  770.0 44.8 48.0 51.2 15.0 10.5
6  950.0 48.3 51.7 55.1 16.2 11.2 6 1250.0 52.0 56.0 59.7 17.9 11.7
6 1600.0 56.0 60.0 64.0 15.0  9.6 6 1550.0 56.0 60.0 64.0 15.0  9.6
6 1650.0 59.0 63.4 68.0 15.9 11.0
7    6.7  9.3  9.8 10.8 16.1  9.7 7    7.5 10.0 10.5 11.6 17.0 10.0
7    7.0 10.1 10.6 11.6 14.9  9.9 7    9.7 10.4 11.0 12.0 18.3 11.5
7    9.8 10.7 11.2 12.4 16.8 10.3 7    8.7 10.8 11.3 12.6 15.7 10.2
7   10.0 11.3 11.8 13.1 16.9  9.8 7    9.9 11.3 11.8 13.1 16.9  8.9
7    9.8 11.4 12.0 13.2 16.7  8.7 7   12.2 11.5 12.2 13.4 15.6 10.4
7   13.4 11.7 12.4 13.5 18.0  9.4 7   12.2 12.1 13.0 13.8 16.5  9.1
7   19.7 13.2 14.3 15.2 18.9 13.6 7   19.9 13.8 15.0 16.2 18.1 11.6
;
run;
```

Dan program untuk melakukan analisis diskriminan kanoniknya adalah :

```
proc candisc data=fish ncan=6 out=outcan;
   class Species;
   var Weight Length1 Length2 Length3 Height Width;
   title 'Analisis Diskriminan Kanonik data Ikan';
   run;
```

Pernyataan CLASS memerintahkan PROC CANDISC menggunakan SPECIES sebagai kelompok. Sedangkan pernyataan VAR menunjukkan peubah mana saja yang digunakan sebagai penjelas.

PROC CANDISC akan menduga fungsi diskriminan kanonik. Fungsi diskriminan yang terbentuk sebanyak peubah penjelas atau banyaknya kelompok dikurangi satu, tergantung mana yang terkecil.

Kali ini kita memiliki 6 peubah penjelas dan 7 kelompok. Sehingga akan ada 6 fungsi diskirminan kanonik.

Fungsi ini berbentuk kombinasi linear dari peubah penjelas, dimana tiap koefisien yang terpilih telah memaksimumkan jarak antar rataan kelompok.

Output dari PROC CANDISC adalah :

```
                              Analisis Diskriminan Kanonik data Ikan

                         The CANDISC Procedure
       Observations     158          DF Total              157
       Variables          6          DF Within Classes     151
       Classes            7          DF Between Classes      6

                        Class Level Information
                Variable
    Species     Name          Frequency       Weight    Proportion

    Bream       Bream                34      34.0000      0.215190
    Parkki      Parkki               11      11.0000      0.069620
    Perch       Perch                56      56.0000      0.354430
    Pike        Pike                 17      17.0000      0.107595
    Roach       Roach                20      20.0000      0.126582
    Smelt       Smelt                14      14.0000      0.088608
    Whitefish   Whitefish             6       6.0000      0.037975
```

PROC CANDISC mendaftar banyaknya amatan di setiap kelompok, serta proporsinya. Kegunaan dari informasi ini akan dibahas kemudian.

Output selanjutnya adalah

```
                      Analisis Diskriminan Kanonik data Ikan

                            The CANDISC Procedure

                Multivariate Statistics and F Approximations
                            S=6    M=-0.5    N=72

Statistic                        Value    F Value    Num DF    Den DF    Pr > F
Wilks' Lambda               0.00036325      90.71        36    643.89    <.0001
Pillai's Trace              3.10465132      26.99        36       906    <.0001
Hotelling-Lawley Trace     52.05799676     209.24        36    413.64    <.0001
Roy's Greatest Root        39.13499776     984.90         6       151    <.0001

          NOTE: F Statistic for Roy's Greatest Root is an upper bound.
```

Uji diatas digunakan untuk menguji hipotesis nol bahwa semua korelasi kanonik adalah nol, yang berarti tidak ada fungsi linear yang bisa membedakan kelompok. Pada output di atas karena nilai p-nya kecil maka hipotesis nol ditolak.

Korelasi kanonik adalah ukuran keeratan hubungan antara peubah penjelas dengan kelompok.

```
                      Analisis Diskriminan Kanonik data Ikan

                              The CANDISC Procedure

                                Adjusted    Approximate        Squared
                  Canonical    Canonical       Standard      Canonical
                Correlation  Correlation          Error    Correlation

              1    0.987463     0.986671       0.001989       0.975084
              2    0.952349     0.950095       0.007425       0.906969
              3    0.838637     0.832518       0.023678       0.703313
              4    0.633094     0.623649       0.047821       0.400809
              5    0.344157     0.334170       0.070356       0.118444
              6    0.005701      .             0.079806       0.000033

                                  Test of H0: The canonical correlations in
                                                    the current row and all
           Eigenvalues of Inv(E)*H                   that follow are zero
           = CanRsq/(1-CanRsq)
                                           Likelihood Approximate
  Eigenvalue Difference Proportion Cumulative      Ratio     F Value Num DF Den DF Pr > F

1    39.1350    29.3859     0.7518     0.7518 0.00036325       90.71     36 643.89 <.0001
2     9.7491     7.3786     0.1873     0.9390 0.01457896       46.46     25 547.58 <.0001
3     2.3706     1.7016     0.0455     0.9846 0.15671134       23.61     16 452.79 <.0001
4     0.6689     0.5346     0.0128     0.9974 0.52820347       12.09      9 362.78 <.0001
5     0.1344     0.1343     0.0026     1.0000 0.88152702        4.88      4    300 0.0008
6     0.0000                0.0000     1.0000 0.99996749        0.00      1    151 0.9442
```

Banyaknya fungsi diskriminan kanonik yang PROC CANDISC hasilkan adalah yang terkecil antara derajat bebas kelompok (grup) dan derajat bebas peubah penjelas. Derajat bebas kelompok adalah banyaknya

kelompok dikurangi satu sedangkan derajat bebas peubah penjelas adalah banyaknya peubah penjelas. Dengan demikian, kita peroleh 6 buah fungsi diskriminan kanonik.

Perhatikan korelasi antara peubah-peubah penjelas dengan fungsi diskriminannya. Korelasi ini adalah korelasi bivariate antara setiap peubah dengan setiap fungsi diskriminan.

```
                    Analisis Diskriminan Kanonik data Ikan

                           The CANDISC Procedure

                         Total Canonical Structure

Variable      Can1       Can2        Can3        Can4         Can5        Can6

Weight      0.230560  0.420828   0.409512   -0.117922  -0.573797   0.508819
Length1     0.102173  0.670793   0.490738   -0.096086  -0.295203   0.449885
Length2     0.119374  0.664922   0.506495   -0.094508  -0.302814   0.431814
Length3     0.222321  0.665900   0.484137   -0.075323  -0.298030   0.422216
Height      0.763233  0.131542   0.484394   -0.070361  -0.197377   0.348750
Width       0.240650  0.273018   0.696271    0.104958  -0.307330   0.526581
```

Matriks Total Canonical Structure memuat nilai korelasi antara peubah penjelas danfungsi diskriminan, tanpa pengkoreksian (penyesuaian) keanggotaan di setiap kelompok, sehingga ada kemungkinan terjadi kerancuan karena hal itu.

```
                        Pooled Within Canonical Structure

Variable        Can1      Can2        Can3        Can4        Can5        Can6

Weight      0.046034   0.162357   0.282143  -0.115459  -0.681453   0.643590
Length1     0.025554   0.324182   0.423532  -0.117850  -0.439169   0.712822
Length2     0.030163   0.324651   0.441629  -0.117107  -0.455129   0.691230
Length3     0.057538   0.333012   0.432369  -0.095598  -0.458798   0.692253
Height      0.243560   0.081113   0.533406  -0.110109  -0.374655   0.705044
Width       0.052710   0.115551   0.526255   0.112738  -0.400405   0.730679
```

Sedangkan matriks Pooled Within Canonical Structure memuat nilai korelasi antara peubah penjelas dan fungsi diskriminan yang telah dikoreksi oleh informasi keanggotaan kelompok.

Dengan melihat lebih jauh terhadap nilai-nilai korelasi pada matriks di atas, variabel yang berkorelasi tinggi dengan fungsi diskriminan kanonik yang pertama adalah HEIGHT, yang menunjukkan bahwa peubah ini erat kaitannya dengan perbedaan spesies. Hal ini berarti bahwa pembeda utama dari spesies ikan-ikan di atas adalah TINGGI (HEIGHT) -nya.

Selanjutnya, untuk fungsi diskriminan kanonk kedua, peubah-peubah yang berkorelasi tinggi adalah LENGTH1 (Panjang Tubuh), LENGTH2 (Panjang Kepala), dan LENGTH3 (Panjang ekor). Peubah yang berhubungan dengan dimensi panjang anatomi inilah yang membentuk fungsi diskriminan kanonik yang kedua.

PROC CANDISC juga menghasilkan koefisien mentah dan koefisien baku penyusun fungsi diskriminan kanonik dari semua peubah, hanyasaja interpretasi terhadapnya cukup sulit. Koefisien mentah dipengaruhi oleh skala dari peubah dan korelasi antar peubah. Sedangkan koefisien baku dipengaruhi oleh korelasi antar peubah penjelas.

Analisis Diskriminan Kanonik data Ikan

The CANDISC Procedure

Total-Sample Standardized Canonical Coefficients

Variable	Can1	Can2	Can3	Can4	Can5	Can6
Weight	-0.23287040	-1.87861652	-2.00951527	-0.32882654	-2.79547487	0.28003904
Length1	-3.30254226	-6.28154797	-29.41614257	-4.51389540	18.02442316	32.99160100
Length2	-26.71741394	-7.41786401	43.47030076	-18.12471382	-13.47860874	-32.79775341
Length3	30.20558626	20.98357040	-13.25763695	21.95619508	-3.19757135	-1.08873008
Height	4.80387980	-3.06026678	1.21255849	-4.18217355	1.49325385	0.72765559
Width	-2.44489997	-1.53319065	1.25337213	4.20224089	-0.07023681	0.69169212

Pooled Within-Class Standardized Canonical Coefficients

Variable	Can1	Can2	Can3	Can4	Can5	Can6
Weight	-0.18772539	-1.51442189	-1.61994419	-0.26507917	-2.25353514	0.22574977
Length1	-2.12531351	-4.04241874	-18.93042395	-2.90486604	11.59941250	21.23136955
Length2	-17.01855083	-4.72505670	27.68986268	-11.54514296	-8.58565086	-20.89162652
Length3	18.78500902	13.04979006	-8.24499242	13.65467033	-1.98858602	-0.67708682
Height	2.42294578	-1.54351499	0.61158138	-2.10937412	0.75315646	0.36700961
Width	-1.79661023	-1.12664977	0.92102794	3.08797458	-0.05161281	0.50828302

Raw Canonical Coefficients

Variable	Can1	Can2	Can3	Can4	Can5	Can6
Weight	-0.000648508	-0.005231659	-0.005596192	-0.000915731	-0.007784969	**0.000779866**
Length1	**-0.329435762**	**-0.626598051**	**-2.934324102**	-0.450270868	1.797975352	3.290983844
Length2	-2.486133674	-0.690253987	4.045038893	-1.686557742	-1.254224048	-3.051927082
Length3	2.595648437	1.803175454	-1.139264914	1.886755746	-0.274776030	-0.093557546
Height	1.121983854	-0.714749340	0.283202557	-0.976779477	0.348761164	0.169949677
Width	-1.446386704	-0.907025481	0.741486686	2.486018007	-0.041551633	0.409200496

Yang perlu juga adalah pembahsan mengenai rata-rata kelompok dari setiap fungsi diskriminan kanonik yang terbentuk. Ruang yang dibentuk oleh fungsi-fungsi ini merupakan ruang yang memaksimumkan jarak antara centroid-centroid kelompok.

Class Means on Canonical Variables

Species	Can1	Can2	Can3	Can4	Can5	Can6
Bream	10.94142464	0.52078394	0.23496708	-0.07271487	-0.17753724	0.00207811
Parkki	2.58903743	-2.54722416	-0.49326158	-1.65865810	0.87343845	-0.00814569
Perch	-4.47181389	-1.70822715	1.29281314	-0.09241115	-0.07247807	0.00237500
Pike	-4.89689441	8.22140791	-0.16469132	-0.48831679	0.01971053	-0.00057099
Roach	-0.35837149	0.08733611	-1.10056438	1.63915159	0.51514591	0.00168629
Smelt	-4.09136653	-2.35805841	-4.03836098	-0.44067181	-0.47039121	0.00070315
Whitefish	-0.39541755	-0.42071778	1.06459242	1.26338797	-0.59421743	-0.02465266

Nampaknya, fungsi diskriminan yang pertama memisahkan ikan BREAM dan PARKKI dengan ikan lainnya. Sedangkan fungsi diskriminan kanonik yang kedua memisahkan ikan PIKE.

Apa yang sudah kita pelajari sejauh ini ?

Nampaknya ada suatu fungsi diskriminan linear dari peubah-peubah penjelas yang secara signifikan memisahkan spesies ikan. Nampaknya fungsi-fungsi diskriminan tersebut bisa diintrepretasikan. Kita masih membutuhkan cara untuk menilai seberapa bagus kemampuan fungsi tersebut memprediksi keanggotaan dari suatu objek berdasarkan peubah yang diamati. Untuk itu kita gunakan PROC DISCRIM. PROC CANDISC berguna untuk menilai derajat atau tingkat kemampuan peubah membedakan antar kelompok.

10.8. Latihan

1. Data berikut merupakan hasil pengukuran empat peubah yang merupakan karakteristik dari dua spesies kutu serangga yaitu haltica oleracea dan haltica carduorum.

Haltica oleracea					*Haltica carduorum*				
No.	y_1	y_2	y_3	y_4	No.	y_1	y_2	y_3	y_4
1	189	245	137	163	1	181	305	184	209
2	192	260	132	217	2	158	237	133	188
3	217	276	141	192	3	184	300	166	231
4	221	299	142	213	4	171	273	162	213
5	171	239	128	158	5	181	297	163	224
6	192	262	147	173	6	181	308	160	223
7	213	278	136	201	7	177	301	166	221
8	192	255	128	185	8	198	308	141	197
9	170	244	128	192	9	180	286	146	214
10	201	276	146	186	10	177	299	171	192
11	195	242	128	192	11	176	317	166	213
12	205	263	147	192	12	192	312	166	209
13	180	252	121	167	13	176	285	141	200
14	192	283	138	183	14	169	287	162	214
15	200	294	138	188	15	164	265	147	192
16	192	277	150	177	16	181	308	157	204
17	200	287	136	173	17	192	276	154	209
18	181	255	146	183	18	181	278	149	235
19	192	287	141	198	19	175	271	140	192
					20	197	303	170	205

   a. Carilah koefisien fungsi diskriminannya
   b. Carilah koefisien bakunya
   c. Hitung uji-t untuk peubah individu

11

11.Analisis Biplot (*Biplot Analysis*)

11.1. Pendahuluan

Dalam Analisis Mutivariat terdapat banyak metode yang dapat digunakan untuk menyelesaikan masalah atau mengolah data yang melibatkan banyak variabel. Misalnya pada saat melakukan suatu penelitian, data yang diperoleh adalah rekapan data yang berupa tabel nilai rata-rata dari beberapa peubah/variabel pada beberapa objek. Semakin banyak peubah yang diukur dan semakin banyak objek yang diamati, maka ukuran tabel yang dimiliki akan semakin besar dan semakin sulit untuk menginterpretasikannya. Untuk itu diperlukan suatu metode yang mampu mempermudah interpretasi dari data yang dimiliki. Metode Mutivariat yang dapat digunakan untuk mengatasi masalah tersebut salah satunya adalah biplot.

Biplot adalah salah satu upaya menggambarkan data -data yang ada pada tabel ringkasan dalam grafik berdimensi dua. Informasi yang diberikan oleh biplot mencakup objek dan peubah dalam satu gambar (Sartono dkk, 2003). Analisis biplot bersifat deskriptif dengan dimensi dua yang dapat menyajikan secara visual segugus objek dan variabel dalam satu grafik. Grafik yang dihasilkan dari biplot ini merupakan grafik yang berbentuk bidang datar. Dengan penyajian seperti ini, ciri-ciri variabel dan objek pengamatan serta posisi relatif antara objek pengamatan

dengan variabel dapat dianalisis (Jollife (1986) dan Rowling (1988) dalam Sartono dkk, 2003).

Tiga hal penting yang bisa didapatkan dari tampilan biplot adalah (Sartono dkk, 2003):

1. *Kedekatan antar objek yang diamati*

Informasi ini dapat dijadikan panduan untuk mengetahui objek yang memiliki kemiripan karakteristik dengan objek lain. Penafsiran ini mungkin akan berbeda untuk setiap bidang terapan, namun inti dari penafsiran ini adalah bahwa dua objek yang memiliki karakteristik sama akan digambarkan sebagai dua titik dengan posisi yang berdekatan.

2. *Keragaman peubah*

Informasi ini digunakan untuk melihat apakah ada variabel yang mempunyai nilai keragaman yang hampir sama untuk setiap objek. Dengan informasi ini, bisa diperkirakan pada variabel mana strategi tertentu harus ditingkatkan, dan juga sebaliknya. Dalam biplot, variabel yang mempunyai nilai keragaman yang kecil digambarkan sebagai vektor pendek sedangkan variable dengan nilai keragaman yang besar digambarkan sebagai vektor yang panjang.

3. *Korelasi antar peubah*

Dari informasi ini bisa diketahui bagaimana suatu variabel mempengaruhi ataupun dipengaruhi variabel yang lain. Pada biplot, variabel akan digambarkan sebagai garis berarah. Dua variabel yang memiliki nilai korelasi positif akan digambarkan sebagai dua buah garis dengan arah yang sama atau membentuk sudut sempit. Sementara itu, dua variabel yang memiliki nilai korelasi negatif akan digambarkan dalam bentuk dua garis dengan arah yang berlawanan atau membentuk sudut lebar (tumpul). Sedangkan dua variabel yang tidak berkorelasi akan

digambarkan dalam bentuk dua garis dengan sudut yang mendekati 90^0 (siku-siku).

4. *Nilai peubah pada suatu objek*

Dalam informasi ini digunakan untuk melihat keunggulan dari setiap objek. Objek yang terletak searah dengan arah vektor variabel dikatakan bahwa objek tersebut mempunyai nilai di atas rata-rata. Namun jika objek terletak berlawanan dengan arah dari vektor variabel tersebut, maka objek tersebut memiliki nilai di bawah rata-rata. Sedangkan objek yang hampir berada ditengah-tengah berarti objek tersebut memiliki nilai dekat dengan rata -rata.

Perlu dipahami sebelumnya bahwa biplot adalah upaya membuat gambar di ruang berdimensi banyak menjadi gambar di ruang berdimensi dua. Pereduksian dimensi ini mengakibatkan menurunnya informasi yang terkandung dalam biplot. Biplot yang mampu memberikan informasi sebesar 70% dari seluruh informasi dianggap cukup.

11.2. Penguraian Nilai Singular (*Singular Value Decomposition*)

Analisis biplot pertama kali diperkenalkan oleh Gabriel (1971). Analisis ini didasarkan pada Singular Value Decomposition (SVD). Biplot dapat dibangun dari suatu matriks data, dengan masing-masing kolom mewakili suatu variabel, dan masing-masing baris mewakili objek penelitian (Udina: 2005).

$$X = \begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1p} \\ X_{21} & X_{22} & \cdots & X_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ X_{n1} & X_{n2} & \cdots & X_{np} \end{bmatrix}$$

Matriks *X* adalah matriks yang memuat variabel-variabel yang akan diteliti sebanyak *p* dan objek penelitian sebanyak *n*. Pendekatan langsung untuk

mendapatkan nilai singularnya, dengan persamaan yang digunakan adalah matriks *X* berukuran *n* x *p* yang berisi *n* objek dan *p* variabel yang dikoreksi terhadap rata-ratanya dan mempunyai rank *r*, dapat dituliskan menjadi

$$_nX_p = {_nU_r}\, {_rL_r}\, {_rA'_p}$$

dengan (r ≤{n,p})

U dan ***A*** adalah matriks dengan kolom ortonormal **($U'U=A'A=I_r$)** dan ***L*** adalah matriks diagonal berukuran *(rxr)* dengan unsur-unsur diagonalnya adalah akar dari nilai eigen–nilai eigen ***X'X***, yaitu $\sqrt{\lambda_1} \geq \sqrt{\lambda_2} \geq \cdots \geq \sqrt{\lambda_r}$. Unsur – unsur diagonal matriks ***L*** ini disebut nilai singular matriks *X* dan kolom-kolom matriks ***A*** adalah vektor eigen dari ***X'X***. Kolom-kolom untuk matriks ***U*** diperoleh dari $u_i = \frac{1}{\sqrt{\lambda_i}} a_i$, i=1,2,….r dengan u_i adalah kolom mariks ***U***, a_i adalah kolom matriks **A** dan λ_i adalah nilai eigen ke-i.

Unsur–unsur diagonal matriks ***L*** merupakan nilai singular dari matriks ***X***. Kemudian didefinisikan L^{α} dengan 0 ≤ a ≤ 1 adalah matriks diagonal berukuran rxr dengan unsur-unsur diagonalnya $\sqrt{\lambda_1{}^{\alpha}} \geq \sqrt{\lambda_2{}^{\alpha}} \geq \cdots \geq \sqrt{\lambda_r{}^{\alpha}}$, dan definisi ini berlaku pula untuk $L^{1-\alpha}$ dengan unsur-unsur digonalnya adalah $\sqrt{\lambda_1{}^{1-\alpha}} \geq \sqrt{\lambda_2{}^{1-\alpha}} \geq \cdots \geq \sqrt{\lambda_r{}^{1-\alpha}}$.

Menurut Jollife (1986),misalkan $\boldsymbol{G} = \boldsymbol{UL^{\alpha}}$ dan $\boldsymbol{H'} = \boldsymbol{L^{1-\alpha}A'}$ dengan a besarnya 0 ≤ a ≤ 1. Persamaan di atas menjadi

$$\boldsymbol{X} = \boldsymbol{UL^{\alpha}L^{1-\alpha}A'} = \boldsymbol{GH'}$$

Hal ini berarti unsur ke-(*i,j*) matriks *X* dapat dituliskan sebagai :

$$x_{ij} = g_i{}'_j$$

dengan g_i', i = 1,2,…, n dan , $_j'$, i = 1,2…, p masing–masing merupakan baris matriks G dan kolom matriks H. Pada g_i dan $_j$ mempunyai *r* dimensi. Jika *X* mempunyai rank dua, vektor baris g_i dan vektor $_j$ dapat digambarkan dalam ruang berdimensi dua. Jika *X* mempunyai rank lebih dua maka persamaan di atas menjadi

$$x_{ij} = \sum_{k=1}^{r} u_{ik} \lambda_k^{\frac{1}{2}} a_{jk}$$

dengan u_{ik} adalah elemen ke-(*i,k*) dari matriks ***U***, a_{jk} adalah elemen ke-(*j,k*) dari matriks A dan $\lambda_k^{\frac{1}{2}}$ adalah elemen diagonal ke-*k* dari matriks ***L.***

Himpunan data asal yang terdiri dari *n* objek dan *p* variabel tereduksi menjadi himpunan data yang terdiri dari *n* objek dengan *m* unsur pertama. Jika ada sebanyak m elemen unsur yang dipertahankan, persamaan di atas dapat didekati dengan

$$_m x_{ij} = \sum_{k=1}^{m} u_{ik} \lambda_k^{\frac{1}{2}} a_{jk}, \quad m < r$$

$$_m x_{ij} = \sum_{k}^{m} u_{ik} \left(\lambda_k^{\alpha}\right)^{\frac{1}{2}} \left(\lambda_k^{1-\alpha}\right)^{\frac{1}{2}} \alpha_{jk}$$

$$= \sum_{k=1}^{m} g_{ikjk}$$

$$= g_i^{*\prime}{}_j^{*}$$

Dengan $g_i^{*\prime}$dan $_j^{*}$masing-masing berisi elemen unsur vektor g_i dan $_j$ Gabriel (1971) menyatakan *m* = 2 disebut biplot, sehingga persamaan yang terakhir dapat dinyatakan sebagai

$$_2 x_{ij} = g_i^{*\prime}{}_j^{*}$$

Dengan ${}_2x_{ij}$ merupakan unsur pendektan matriks ***X*** pada dimensi dua, sedangkan $g_i{}^{*\prime}$ dan ${}_j{}^*$ masing –masing mengandung dua unsur pertama vektor g_i dan ${}_j$.

Dari pendekatan matriks *X* pada dimensi dua diperoleh matriks **G** dan **H** sebagai berikut

$$G = \begin{bmatrix} g_{11} & g_{12} \\ \vdots & \vdots \\ g_{i1} & g_{i2} \\ \vdots & \vdots \\ g_{n1} & g_{n2} \end{bmatrix} \textit{dan } H = \begin{bmatrix} \boldsymbol{h}_{11} & \boldsymbol{h}_{12} \\ \vdots & \vdots \\ \mathbf{h}_{i1} & \mathbf{h}_{i2} \\ \vdots & \vdots \\ \mathbf{h}_{p1} & \mathbf{h}_{p2} \end{bmatrix}$$

Matriks **G** adalah titik-titik koordinat dari n objek dan matriks H adalah titik-titik koordinat dari *p* variabel.

Gabriel (1971) mengemukakan ukuran pendekatan matriks X dengan biplot dalam bentuk:

$$\rho^2 = \frac{(\lambda_1 + \lambda_2)}{\sum_{k=1}^{r} \lambda_k}$$

dengan λ_1 adalah nilai eigen terbesar ke-1, λ_2 adalah nilai eigen terbesar ke-2 dan λ_k ,*k* =1,2….r adalah nilai eigen ke-k. Apabila ρ^2 mendekati nilai satu, maka biplot memberikan penyajian yang semakin baik mengenai informasi data yang sebenarnya .

Menurut Jollife (1986) untuk mendeskripsikan biplot perlu mengambil nilai a dalam mendefinisikan **G** dan **H**.Pemilihan nilai a pada $\boldsymbol{G} = \boldsymbol{UL^{\alpha}}$ dan $\boldsymbol{H'} = \boldsymbol{L^{1-\alpha}A'}$ bersifat sembarang dengan syarat 0 ≤a≤ 1. Pengambilan nilai ekstrim a=0 dan a=1 berguna dalam interpretasi biplot.

Jika a=0 didapat $\boldsymbol{G} = \boldsymbol{UL^0} = \boldsymbol{U}$ dan $\boldsymbol{H'} = \boldsymbol{L^1A'}$ **=LA'** sehingga

$$X'X = (GH')'(GH')$$

$$= HG'GH'$$

$$= HU'UH'$$

$$= HH^{\wedge'}$$

Matriks **U** ortonormal dan **X'X** = **(n-1)S** dengan n adalah banyaknya objek pengamatan dan **S** adalah matriks kovarian dari matriks **X** maka

HH'=(n-1) S

Hasil kali $h_i\ h_j$ adalah akan sama dengan *(n-1)* kali kovarian s_{jk} antara variabel ke-j dan variabel ke-k. Selanjutnya untuk mengetahui variansi variabel digunakan matriks H.

$$HH' = \begin{bmatrix} h_{11} & h_{12} \\ \dots & \dots \\ h_{p1} & h_{p2} \end{bmatrix} \begin{bmatrix} h_{11} & \dots & h_{p1} \\ h_{12} & \dots & h_{p2} \end{bmatrix}$$

$$= \begin{bmatrix} h_{11}^2 + h_{12}^2 & \dots & h_{11}h_{p1} + h_{12}h_{p2} \\ \dots & \dots & \dots \\ h_{11}h_{p1} + h_{12}h_{p2} & \dots & h_{p1}^2 + h_{p2}^2 \end{bmatrix}$$

Diagonal utama pada matriks ***HH'***: $h_{11}^2 + h_{12}^2$, … , $h_{j1}^2 + h_{j2}^2$, …, $h_{p1}^2 + h_{p2}^2$ menggambarkan variansi dari variabel. Sedangkan $h_{j1}^2 + h_{j2}^2$, j= 1,2,…,*n* menyatakan panjang vector variabel (dengan jarak Euclid dari titik O(0,0)). Sehingga dapat disimpulkan bahwa panjang vector variabel sebanding dengan variansi variabel.

Nilai cosinus sudut antara dua vektor peubah menggambarkan korelasi kedua peubah. Semakin sempit sudut yang dibuat antara dua variabel

maka semakin tingggi korelasinya. Korelasi peubah ke–*j* dan ke-*k* sama dengan nilai cosinus sudut vektor $\boldsymbol{h_j}$ dan $\boldsymbol{h_k}$.

$$\boldsymbol{S}_{j\cdot k} = |\boldsymbol{h}_j||\boldsymbol{h}_k| \cos\theta$$

$$\cos\theta = \frac{S_{j\cdot k}}{|\boldsymbol{h}_j||\boldsymbol{h}_k|} = \frac{S_{jk}}{\sqrt{S_{jj}}\sqrt{S_{kk}}} = \frac{S_{jk}}{S_j S_k} = r_{jk}$$

Kedekatan antar obyek pada gambar biplot dapat dilihat dengan menggunakan jarak Euclid antara $\boldsymbol{g_i}$ dan $\boldsymbol{g_j}$ sebanding dengan jarak Mahalanobis antar objek pengamatan $\boldsymbol{x_i}$ dan $\boldsymbol{x_j}$ dalam data pengamatan sesungguhnya.

Jarak Mahalanobis antara dua pengamatan $\boldsymbol{x_i}$ dan $\boldsymbol{x_j}$ didefinisikan sebagai :

$$\delta^2(x_i, x_j) = (x_i - x_j)' S^{-1}(x_i - x_j)$$

Jarak Euclid antara dua pengamatan $\boldsymbol{g_i}$ dan $\boldsymbol{g_j}$ didefinisikan sebagai :

$$d^2(g_i, g_j) = (g_i - g_j)'(g_i - g_j)$$

Menurut Jollife (1986) $\delta^2(x_i, x_j) = (n-1)^2 d^2(g_i, g_j)$. Hal ini dapat dibuktikan sebagai berikut: Persamaan yang ketiga di atas dapat ditulis kembali sebagai $x_i' = g_i H'$, $i = 1, 2, \ldots, n$ dan disubsitusikan ke dalam persamaan sehingga menghasilkan:

$$\begin{aligned}
\delta^2(x_i, x_j) &= (Hg_i - Hg_j)' S^{-1}(Hg_i - Hg_j) \\
&= (H(g_i - g_j))' S^{-1} H(g_i - g_j) \\
&= (g_i - g_j)' H S^{-1} H(g_i - g_j) \\
&= (n-1)(g_i - g_j)'(LA)'(X'X)^{-1}(LA)(g_i - g_j)
\end{aligned}$$

Dengan H' = $LA^{'}((\alpha = 0)$ dan $S^{-1} = (n-1)(X'X)^{-1}$

Sedangkan

$$
\begin{aligned}
X'X &= (ULA')'(ULA)' \\
&= ALU'ULA' \\
&= AL^2A'
\end{aligned}
$$

dan

$$
\begin{aligned}
(X'X)^{-1} &= ((ULA')'(ULA'))^{-1} \\
&= (U'LAULA')^{-1} \\
&= (ALU'ULA')^{-1} \\
&= AL^{-2}A'
\end{aligned}
$$

Subsitusikan persamaan (12) dan (13) ke persamaan (11) menghasilkan

$$
\begin{aligned}
\delta^2(x_i, x_j) &= (n-1)^2(g_i, g_j)'(LA')(X'X)^{-1}LA(g_i, g_j) \\
&= (n-1)^2(g_i, g_j)'L(A'A)L^{-2}(A'A)L(g_i, g_j) \\
&= (n-1)^2(g_i, g_j)'LL^{-2}L(g_i, g_j), \text{ (A adalah ortogonal)} \\
&= (n-1)^2(g_i, g_j)'(g_i, g_j) \\
&= (n-1)^2 d^2(g_i, g_j).
\end{aligned}
$$

Berarti dapat dilihat bahwa Mahalanobis sebanding dengan jarak Euclid. Hal ini menunjukkan bahwa jarak Euclid mampu menggambarkan posisi objek pengamatan dalam data pengamatan sesungguhnya. Jika a=1 maka ***G=UL*** dan ***H=A*** sehingga diperoleh :

XX' = (GH')(GH')'

= GH'HG'

= GA'AG'

= GG'

Pada keadaan ini, jarak Euclid antara $\boldsymbol{g_i}$ dan $\boldsymbol{g_j}$ akan sama dengan jarak Euclid antara objek pengamatan $\boldsymbol{x_i}$ dan $\boldsymbol{x_j}$.Vektor baris ke-i sama dengan skor komponen utama untuk responden ke-i dari hasil analisis komponen utama. Untuk **G = UL** maka unsur ke-*k* dari $\boldsymbol{g_i}$ adalah $\boldsymbol{u_{ik}}\sqrt{\boldsymbol{\lambda_k}}$. Hasil tersebut sama dengan $\boldsymbol{Z_{ik}}$ yang merupakan skor komponen utama ke-k dari objek ke-i. Sedangkan **H=A** diperoleh bahwa vektor pengaruh kolom $\boldsymbol{h_j}$ sama dengan $\boldsymbol{\alpha_j}$.

11.3. Ilustrasi

Data yang digunakan merupakan hasil pengamatan mahasiswa holtikultura Jurusan Budi Daya pertanian IPB (Nitrisari, 2002). Tidak semua variabel digunakan, karena mengakibatkan matriks X`X yang tidak definit positif.

Peubah pengamatan buah nenas queen disajikan pada tabel berikut:

No	Peubah	Uraian	Satuan
1	X_1	*Kadar Air*	*%*
2	X_2	*Vitamin C*	*Mg/100gr*
3	X_3	*Suckers*	*Tangkai*
4	X_4	*Shoots*	*Tangkai*
5	X_5	*Jumlah daun*	*Helai*

Analisis Biplot dilakukan sebanyak dua kali, untuk mengetahui posisi relatif antara lokasi tanam dengan setiap peubah pengamatan, dengan cara mengamati posisi vektor yang mewakili peubah, ragam serta korelasi

antar peubah. Analisis Biplot pertama dilakukan untuk melihat posisi relatif antara lokasi tanam dengan peubah kualitas buah yang menyangkut kandungan kimia dan morfologi buah. Analisis Biplot yang kedua dilakukan untuk melihat posisi relatif antara lokasi tanam dengan setiap peubah morfologi tanaman.

Data Kualitas Buah dan morfologi Tanaman Nenas

Lokasi	X_1	X_2	X_3	X_4	X_5
Cipelang	*84.09*	*9.014*	*2.5*	*3*	*28*
Sukaharja	*81.54*	*9.008*	*1.5*	*6*	*38.5*
Tajur Halang	*83.58*	*9.55*	*2*	*4*	*36.5*
Sukaluyu	*83.79*	*8.263*	*1*	*9*	*31*

Tahapan membuat biplot:

1. Cari nilai X'X

$$X`X=\begin{bmatrix} 27726.2802 & 2983.04535 & 583.485 & 1829.94 & 11141.97 \\ 2983.04535 & 321.875929 & 63.41 & 193.657 & 1203.928 \\ 583.485 & 63.41 & 13.5 & 33.5 & 231.75 \\ 1829.94 & 193.657 & 33.5 & 142 & 740 \\ 11141.97 & 1203.928 & 231.75 & 740 & 4559.5 \end{bmatrix}$$

2. Cari nilai eigen dari X'X

Nilai eigen yang diperoleh adalah 0.1108921790 10-5, 0.02160697849, 22.59176143, 70.93554727, dan 32669.6072. Matriks L diperoleh dengan menggunakan rumus $l_i = \sqrt{\lambda_i}$ dimana l_i merupakan nilai diagonal ke-i dan $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_5$

$$l_1 = \sqrt{32669.6072} = 180.7435$$

$$l_2 = \sqrt{70.93554727} = 8.4223$$
$$l_3 = \sqrt{22.59176143} = 4.753$$
$$l_4 = \sqrt{0.02160697849} = 0.14699$$
$$l_5 = \sqrt{0.1108921790\ 10^{-5}} = 0.001053$$

Dengan menggunakan program SAS 6.12 dapat diperoleh matrik U, L, dan A. Output yang diperoleh adalah

$$U = \begin{bmatrix} 0.4922259 & -0.664929 & -0.464025 & -0.316644 & 0 \\ 0.501689 & 0.6248001 & 0.0440133 & -0.596654 & 0 \\ 0.5076557 & 0.3002489 & -0.378514 & 0.7133468 & 0 \\ 0.4983046 & -0.278109 & 0.7996693 & 0.1867541 & 0 \end{bmatrix}$$

$$L = \begin{bmatrix} 180.74736 & 0 & 0 & 0 & 0 \\ 0 & 8.4223 & 0 & 0 & 0 \\ 0 & 0 & 4.753 & 0 & 0 \\ 0 & 0 & 0 & 0.14699 & 0 \\ 0 & 0 & 0 & 0 & 0.00105 \end{bmatrix}$$

$$A = \begin{bmatrix} 0.9210746 & -0.377051 & -0.013237 & -0.055054 & -0.079044 \\ 0.0991535 & 0.0242083 & -0.166921 & 0.8621129 & 0.4674197 \\ 0.0193459 & -0.047817 & -0.221204 & -0.497593 & 0.8371433 \\ 0.0608705 & 0.0536703 & 0.9583207 & 0.0293804 & 0.2723463 \\ 0.3710938 & 0.9230815 & -0.068207 & -0.072582 & -0.017015 \end{bmatrix}$$

3. Cari nilai G dan H

Apabila digunakan a = 0, maka G = U dan H' = LA'. Matrik G dan H' yang diperoleh adalah sebagai berikut:

$$G = \begin{bmatrix} 0.4922259 & -0.664929 & -0.464025 & -0.316644 & 0 \\ 0.501689 & 0.6248001 & 0.0440133 & -0.596654 & 0 \\ 0.5076557 & 0.3002489 & -0.378514 & 0.7133468 & 0 \\ 0.4983046 & -0.278109 & 0.7996693 & 0.1867541 & 0 \end{bmatrix}$$

$$H` = \begin{bmatrix} 166.4818 & 17.92173 & 3.49672 & 11.00218 & 67.07422 \\ -3.17565 & 0.20389 & -0.40273 & 0.452029 & 7.774492 \\ -0.06292 & -0.79339 & -1.0514 & 4.554973 & -0.32419 \\ -0.00809 & 0.126725 & -0.07314 & 0.004319 & -0.01067 \\ -8.3\ 10^{-5} & 0.000491 & 0.000879 & 0.000286 & -1.8\ 10^{-5} \end{bmatrix}$$

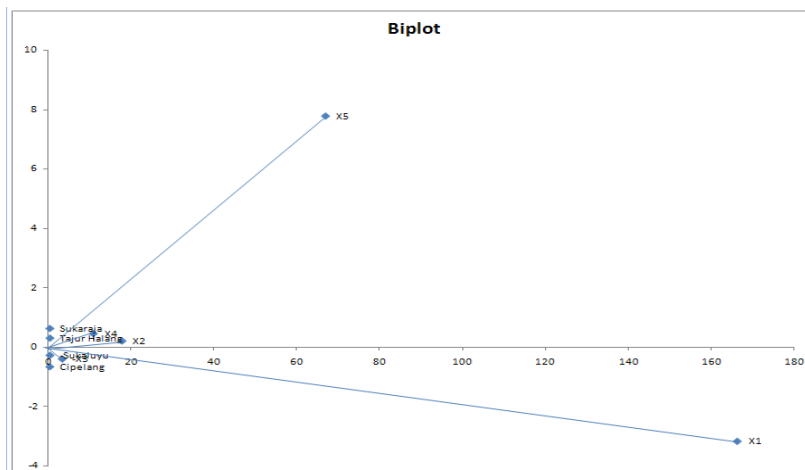
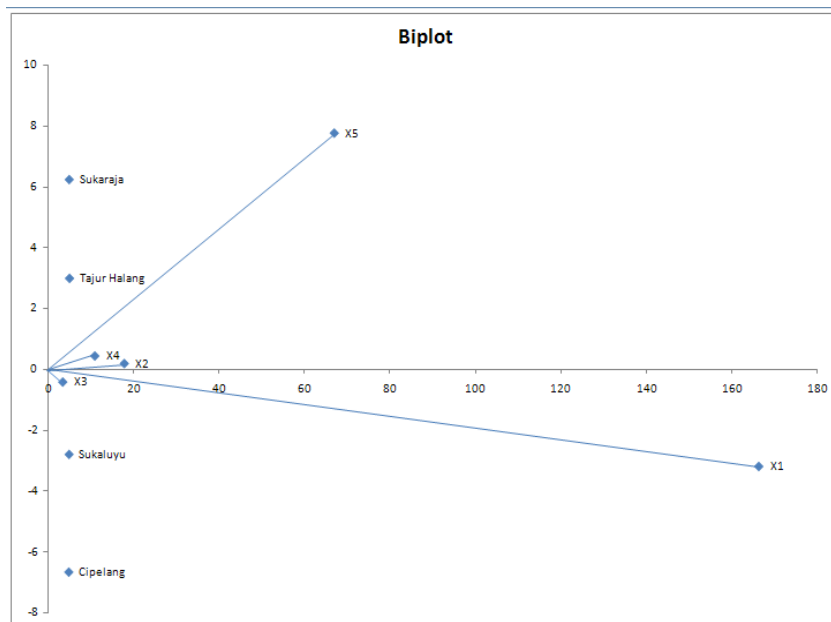
Dari pendekatan matriks X pada dimensi dua, matriks G dan H yang diperoleh adalah sebagai berikut:

$$G^* = \begin{bmatrix} 0.492 & -0.665 \\ 0.502 & 0.625 \\ 0.508 & 0.300 \\ 0.498 & -0.278 \end{bmatrix}$$

$$H^{*`} = \begin{bmatrix} 166.4818 & 17.92173 & 3.49672 & 11.00218 & 67.07422 \\ -3.17565 & 0.20389 & -0.40273 & 0.452029 & 7.774492 \end{bmatrix}$$

Matriks G merupakan titik koordinat dari empat objek dan matrik H merupakan titik koordinat dari 5 variabel.

4. Gambar biplotnya

Berdasarkan nilai dari matriks G dan H selanjutnya digunakan untuk membuat grafik Biplot. Grafik yang diperoleh seperti berikut ini.

Dari gambar biplot tersebut, kita tidak dapat secara jelas melihat sebaran data di objek. Untuk mempermudah melihat sebaran data objek, nilai koordinat objek dikalikan dengan konstanta, misalnya 10, sehingga diperoleh gambar seperti berikut ini.

Dari grafik dapat dilihal bahwa posisi relative dari Sukaraja berdekatan dengan Tajur Halang. Hal ini dapat diartikan bahwa karakteristik dari buah nanas yang dihasilkan di Sukaraja relative sama dengan buah nanas yang dihasilkan di Tajur Halang. Karakteristik dari buah nanas di Sukaraja dan Tajur Halang dicirikan oleh variabel X5, dalam hal ini jumlah daun. Sedangkan hasil buah nanas di Sukaluyu dan Cipelang kuat dicirikan oleh variabel X1, yaitu kadar airnya.

5. Menghitung ukuran kebaikan biplot dan karakteristik lainnya

Ukuran kebaikan biplot

$$\rho^2 = \frac{(\lambda_1 + \lambda_2)}{\sum_{k=1}^{14} \lambda_k} = \frac{(32669.6072 + 70.93554727)}{0.1108921790\ 10^{-5} + \ ... + 32669.6072} = \frac{32740.54}{32763.156} = 0.9993$$

Karena nilai ρ^2 yang diperoleh mendekati nilai 1, berarti biplot yang dihasilkan sangat baik. Informasi yang diberikan oleh biplot sebesar 99.93 % dari keseluruhan informasi yang terkandung dalam data.

Matriks varian covarian dari variabel

$$S = H`H$$

$$= \begin{bmatrix} 32669.61 & -9.84784\ 10^{-5} & -0.00016 & -1.18417\ 10^{-6} & -2.5895\ 10^{-8} \\ -9.8\ 10^{-5} & 70.93554435 & 8.07\ 10^{-6} & -4.81095\ 10^{-7} & 4.4477\ 10^{-10} \\ -0.00016 & 8.0746610^{-6} & 22.59175 & 7.23639\ 10^{-9} & 1.66748\ 10^{-9} \\ -1.2\ 10^{-6} & -4.81095\ 10^{-7} & 7.24\ 10^{-9} & 0.021606995 & 3.41631\ 10^{-11} \\ -2.6\ 10^{-8} & 4.4477\ 10^{-10} & 1.67\ 10^{-9} & 3.41631\ 10^{-11} & 1.1025\ 10^{-6} \end{bmatrix}$$

Dengan menggunakan matriks varian kovarian, dapat dihitung besar korelasi antar variabel dengan menggunakan rumus: $r_{jk} = \frac{S_{jk}}{S_j S_k}$

$$r_{12} = \frac{S_{12}}{S_1 S_2} = \frac{-9.8\ 10^{-5}}{\sqrt{32669.61}\sqrt{70.935}} = -6.4376\ 10^{-8}$$

r_{13}= -1.862 10^{-7}
r_{14} = -4.4567 10^{-8}
r_{15} = -1.3642 10^{-4}
r_{23} = 2.0159 10^{-7}
r_{24} = -3.8857 10^{-7}
r_{25} = 5.0294 10^{-5}
r_{34} = 1.0357 10^{-8}
r_{35} = 3.34 10^{-4}
r_{45}= 2.213 10^{-4}

Nilai korelasi antar variabel sangat kecil, sehingga hubungan yang dimiliki juga sangat lemah. Atau bisa dikatakan kelima variabel tersebut saling bebas.

Jarak Euclid

$$GG` = \begin{bmatrix} 1 & 2.51\ 10^{-7} & 5.77\ 10^{-8} & 5.73\ 10^{-8} \\ 2.51\ 10^{-7} & 1 & -4.86\ 10^{-8} & -9.05\ 10^{-8} \\ 5.77\ 10^{-8} & -4.86\ 10^{-8} & 1 & -3.37\ 10^{-7} \\ 5.73\ 10^{-8} & -9.05\ 10^{-8} & -3.37\ 10^{-7} & 1 \end{bmatrix}$$

Jarak objek Cipelang dengan Sukaharja sebesar $2.51\ 10^{-7}$, jarak objek Cipelang dengan Tajur halang sebesar $5.77\ 10^{-8}$, jarak objek Cipelang dengan objek Sukaluyu sebesar $5.73\ 10^{-8}$, jarak objek Sukaharja dengan objek Tajur halang sebesar $4.86\ 10^{-8}$, dan jarak objek Tajur halang dengan objek Sukaluyu sebesar $3.37\ 10^{-7}$.

11.4. Aplikasi SAS

Untuk analisis biplot, SAS tidak menyediakan suatu prosedur tersendiri namun melalui program makro %BIPLOT yang telah tersedia dalam SAS.

Analisis biplot dapat diterapkan pada data yang minimal memiliki skala pengukuran interval. Namun demikian penerapan untuk data ordinal telah cukup memuaskan. Artinya mampu mewakili keadaan sebenarnya.

Sebagai input untuk makro %BIPLOT adalah matriks rataan, yaitu matriks yang berisi rataan dari setiap peubah pada setiap objek, atau matriks data dari n objek dan p peubah itu sendiri.

Sebagai ilustrasi sebagai berikut:

Dilakukan survey untuk mengetahui karakteristik teh yang disukai oleh konsumen. Terdapat 6 merk teh dan responden diminta untuk memberikan penilaian terhadap 9 atribut untuk masing-masing merk teh tersebut.

Rata-rata nilai untuk masing-masing atribut pada masing-masing merk teh sebagai berikut :

	sosro	Javati 6138	Kepala jenggot	Peko super	Pucuk bola	Javati
Ukuran kemasan	1.3115	-.2131	.9508	-.2397	.1803	0.0165
Desain kemasan	.6803	-.2787	.5328	-.4463	-.3033	-0.0492
Harga	1.6557	-.9098	1.2377	-.6803	-.7541	-.7377
Merk	1.3279	-.2975	.8607	-.1639	-0.0902	-.2869
Warna seduhan	1.1750	0.0248	.8595	0.0749	0.0413	0.0246
Variasi rasa	1.2623	0.0738	.8197	-.1066	-0.0328	.1639
Khasiat	1.4918	-.1721	.6967	-.1393	-0.0902	0.0820
Aroma	1.5164	-.2213	.5574	-.1885	-.1074	-0.0328
Kepraktisan	1.1885	-0.041	.8607	-.1803	-.1230	.1967

Sebaiknya penulisan atribut dan merk tidak terlalu panjang sehingga mudah dilihat pada biplot.

Program macro Biplot sudah tersedia di SAS, tinggal dijalankan saja. Namun demikian setiap akan menjalankan macro Biplot ini maka macro Biplot harus ada di window program editor. Kalau tidak ada maka macro biplot tidak akan jalan. Untuk input data dan perintah *running* diletakkan di bawah program macro biplot.

Program macro biplot tersebut sebagai berikut :

```
/*-------------------------------------------------------------------*
  *    Name: BIPLOT.SAS                                               *
  *   Title: Construct a biplot of observations and variables         *
  *          Uses IML.                                                *
  *     Doc: http://www.math.yorku.ca/SCS/sssg/biplot.html            *
  *-------------------------------------------------------------------*
  *  Author:  Michael Friendly            <friendly@yorku.ca>         *
  * Created:   1 Mar 1989 13:16:36                                    *
  * Revised:  22 Jul 1998 11:08:14                                    *
  * Version:  1.6                                                     *
  * 1.5  Added dimension labels, fixed problem with dim=3,            *
  *      Added colors option, Fixed problem with var=_NUM_            *
  * 1.6  Added power transformation (for log(freq))                   *
  *      Added point symbols, marker styles (interp=)                 *
  *      Made ID optional, can be char or numeric                     *
  *      Fixed bug introduced with ID                                 *
  *                                                                   *
  *      From ``SAS System for Statistical Graphics, First Edition''  *
  *      Copyright(c) 1991 by SAS Institute Inc., Cary, NC, USA       *
  *-------------------------------------------------------------------*/

%macro BIPLOT(
       data=_LAST_,       /* Data set for biplot                        */
       var =_NUM_,        /* Variables for biplot                       */
       id  =ID,           /* Observation ID variable                    */
       dim =2,            /* Number of biplot dimensions                */
       factype=SYM,       /* Biplot factor type: GH, SYM, or JK         */
```

```
        scale=1,           /* Scale factor for variable vectors        */
        power=1,           /* Power transform of response              */
        out =BIPLOT,       /* Output dataset: biplot coordinates       */
        anno=BIANNO,       /* Output dataset: annotate labels          */
        xanno=dim1,
        yanno=dim2,
        zanno=dim3,
        std=MEAN,          /* How to standardize columns: NONE|MEAN|STD*/
        colors=BLUE RED,   /* Colors for OBS and VARS                  */
        symbols=none none, /* Symbols for OBS and VARS                 */
        interp=none vec,   /* Markers/interpolation for OBS and VARS   */
        pplot=NO,          /* Produce printer plot?                    */
        gplot=YES,
        haxis=,            /* AXIS statement for horizontal axis       */
        vaxis=,            /* and for vertical axis- use to equate axes */
        name=biplot);

%let std=%upcase(&std);
%let factype=%upcase(&factype);
      %if &factype=GH  %then %let p=0;
%else %if &factype=SYM %then %let p=.5;
%else %if &factype=JK  %then %let p=1;
%else %do;
   %put BIPLOT: FACTYPE must be GH, SYM, or JK. "&factype" is not valid.;
   %goto done;
   %end;
%if %upcase("&var") ^= "_NUM_" %then %let var={&var};
%if &data=_LAST_ %then %let data=&syslast;

proc iml;
start biplot(y,id,vars,out, g, scale);
   N = nrow(Y);
   P = ncol(Y);
   %if &std = NONE
       %then Y = Y - Y[:] %str(;);             /* remove grand mean */
       %else Y = Y - J(N,1,1)*Y[:,] %str(;);   /* remove column means */
   %if &std = STD %then %do;
      S = sqrt(Y[##,] / (N-1));
      Y = Y * diag (1 / S );
   %end;

   *-- Singular value decomposition:
        Y is expressed as U diag(Q) V prime
        Q contains singular values, in descending order;
    call svd(u,q,v,y);

   reset fw=8 noname;
   percent = 100*q##2 / q[##];
   cum = cusum(percent);
   c1={'Singular Values'};
   c2={'Percent'};
   c3={'Cum % '};
   Print "Singular values and variance accounted for",,
         q       [colname=c1 format=9.4 ]
         percent [colname=c2 format=8.2 ]
         cum     [colname=c3 format=8.2 ];

   d = &dim ;
        *-- Assign macro variables for dimension labels;
        lab     =     '%let     p'     +     char(t(1:d),1)     +     '='     +
left(char(percent[t(1:d)],8,1)) + ';';
        call execute(lab);
        /*
        call execute('%let p1=', char(percent[1],8,1), ';');
        call execute('%let p2=', char(percent[2],8,1), ';');
```

```
        if d > 2 then
        call execute('%let p3=', char(percent[3],8,1), ';');
        */
   *-- Extract first  d  columns of U & V, and first  d  elements of Q;
   U = U[,1:d];
   V = V[,1:d];
   Q = Q[1:d];

   *-- Scale the vectors by QL, QR;
   * Scale factor 'scale' allows expanding or contracting the variable
     vectors to plot in the same space as the observations;
   QL= diag(Q ## g );
   QR= diag(Q ## (1-g));
   A = U * QL;
   B = V * QR;
        ratio = max(sqrt(A[,##])) /  max(sqrt(B[,##]));
        print 'OBS / VARS ratio:' ratio 'Scale:' scale;
        if scale=0 then scale=ratio;
   B = B # scale;
   OUT=A // B;

   *-- Create observation labels;
   id = id // vars`;
   type = repeat({"OBS "},n,1) // repeat({"VAR "},p,1);
   id  = concat(type, id);

   factype = {"GH" "Symmetric" "JK"}[1 + 2#g];
   print "Biplot Factor Type", factype;

   cvar = concat(shape({"DIM"},1,d), char(1:d,1.));
   print "Biplot coordinates",
         out[rowname=id colname=cvar f=9.4];
   %if &pplot = YES %then %do;
   call   pgraf(out[,{1   2}],substr(id,5),'Dimension   1',   'Dimension   2',
'Biplot');
      %end;
   create &out  from out[rowname=id colname=cvar];
   append from out[rowname=id];
finish;

start power(x, pow);
                if pow=1 then return(x);
                if any(x <= 0) then x = x + ceil(min(x)+.5);
                if abs(pow)<.001 then xt =  log(x);
                        else xt = ((x##pow)-1) / pow;
        return (xt);
        finish;

        /*--- Main routine */

   use &data;
   read all var &var into y[ c=vars ];
   %if &id = %str() %then %do;
      id=compress(char(1:nrow(xy),4))`;
   %end;
   %else %do;
      read all var{&id} into id;
   %end;
*   read all var &var into y[colname=vars rowname=&id];

        %if &power ^= 1 %then %do;
                y = power(y, &power);
                %end;

   scale = &scale;
```

```
   run biplot(y, id,vars,out, &p, scale );
   quit;
 /*----------------------------------*
  |  Split ID into _TYPE_ and _NAME_ |
  *----------------------------------*/

data &out;
   set &out;
   drop id;
   length _type_ $3 _name_ $16;
   _type_ = substr(id,1,3);
   _name_ = substr(id,5);
   label
   %do i=1 %to &dim;
       dim&i = "Dimension &i (&&p&i%str(%%))"
       %end;
   ;
 /*--------------------------------------------------*
  | Annotate observation labels and variable vectors |
  *--------------------------------------------------*/
        %*-- Assign colors and symbols;
        %let c1= %scan(&colors,1);
        %let c2= %scan(&colors,2);
        %if &c2=%str() %then %let c2=&c1;

        %let v1= %upcase(%scan(&symbols,1));
        %let v2= %upcase(%scan(&symbols,2));
        %if &v2=%str() %then %let v2=&v1;

        %let i1= %upcase(%scan(&interp,1));
        %let i2= %upcase(%scan(&interp,2));
        %if &i2=%str() %then %let i2=&i1;
data &anno;
   set &out;
   length function color $8 text $16;
   xsys='2'; ysys='2'; %if &dim > 2 %then %str(zsys='2';);
   text = _name_;

   if _type_ = 'OBS' then do;         /* Label observations (row points) */
      color="&c1";
                if "&i1" = 'VEC' then link vec;
      x = &xanno; y = &yanno;
      %if &dim > 2 %then %str(z = &zanno;);
                %if &v1=NONE %then
        %str(position='5';);
                %else %do;
      if dim1 >=0
         then position='>';             /* rt justify         */
         else position='<';             /* lt justify         */
                        %end;
      function='LABEL   '; output;
      end;

   if _type_ = 'VAR' then do;           /* Label variables (col points) */
      color="&c2";
                if "&i2" = 'VEC' then link vec;
      x = &xanno; y = &yanno;
      if dim1 >=0
         then position='6';             /* down justify       */
         else position='2';             /* up justify         */
      function='LABEL   '; output;      /* variable name      */
      end;
                return;

vec:                    /* Draw line from the origin to point */
```

```
      x = 0; y = 0;
      %if &dim > 2 %then %str(z = 0;);
      function='MOVE'    ; output;
      x = &xanno; y = &yanno;
      %if &dim > 2 %then %str(z = &zanno;);
      function='DRAW'    ; output;
                return;
%if &gplot = YES %then %do;
        %if &i1=VEC %then %let i1=NONE;
        %if &i2=VEC %then %let i2=NONE;
        %let legend=nolegend;

        %let warn=0;
   %if %length(&haxis)=0 %then %do;
                %let warn=1;
                axis2 offset=(1,5) ;
      %let haxis=axis2;
      %end;
   %if %length(&vaxis)=0 %then %do;
                %let warn=1;
                axis1 offset=(1,5) label=(a=90 r=0);
      %let vaxis=axis1;
      %end;
        proc gplot data=&out &GOUT;
                plot dim2 * dim1 = _type_/
                        anno=&anno frame &legend
                        href=0 vref=0 lvref=3 lhref=3
                        vaxis=&vaxis haxis=&haxis
                        vminor=1 hminor=1
                        name="&name" des="Biplot of &data";
                symbol1 v=&v1 c=&c1 i=&i1;
                symbol2 v=&v2 c=&c2 i=&i2;
        run; quit;
        %if &warn %then %do;
                %put WARNING:  No VAXIS= or HAXIS= parameter was specified,
so the biplot axes have not;
                %put WARNING:   been equated.   This may lead to incorrect
interpretation of distance and;
                %put WARNING:  angles.  See the documentation.;
                %end;
        goptions reset=symbol;
        %end;  /* %if &gplot=YES */

%done:
%mend BIPLOT;
```

Input data untuk %BIPLOT berupa **matriks rataan yang berukuran n atribut x p peubah.**

```
data teh;
input id$ ukuran desain harga merk warna variasi khasiat aroma praktis;
cards;
Sosro     1.3115  0.6803  1.6557  1.3279  1.175   1.2623   1.4918  1.5164   1.1885
Javati6  -0.2131 -0.2787 -0.9098 -0.2975  0.0248  0.0738  -0.1721 -0.2213  -0.041
Kjenggot  0.9508  0.5328  1.2377  0.8607  0.8595  0.8197   0.6967  0.5574   0.8607
PekoSuper-0.2397 -0.4463 -0.6803 -0.1639  0.0749 -0.1066  -0.1393 -0.1885  -0.1803
PucukBola 0.1803 -0.3033 -0.7541 -0.0902  0.0413 -0.0328  -0.0902 -0.1074  -0.123
Javati    0.0165 -0.0492 -0.7377 -0.2869  0.0246  0.1639   0.082  -0.0328   0.1967
;
```

Untuk definisi peubah pada input, peubah yang berisi atribut maka nama peubah harus id$. Sedangkan nama peubah untuk atribut lainnya bebas.

Syntaks untuk menjalankan program makro %BIPLOT sebagai berikut :

```
%biplot;

run;
```

Output %BIPLOT berupa nilai singular dan keragamannya, ratio skala garis pada biplot, koordinat biplot, serta biplot itu sendiri.

Dua nilai singular pertama menunjukkan keragaman yang diterangkan oleh komponen 1 (sumbu utama 1) dan komponen 2 (sumbu utama 2) pada biplot. Besarnya keragaman yang dapat diterangkan oleh kedua sumbu utama tersebut dilihat dari persentase keragamannnya.

Sedangkan interpretasi dari biplot adalah :

1. Panjang vektor peubah sebanding dengan keragaman peubah tersebut. Semakin panjang vektor suatu peubah maka keragaman peubah tersebut semakin tinggi.
2. Nilai cosinus sudut antara dua vektor peubah menggambarkan korelasi kedua peubah. Semakin sempit sudut yang dibuat antara dua peubah maka semakin positif tinggi korelasinya. Jika sudut yang dibuat tegak lurus maka korelasi keduanya rendah. Sedangkan jika sudutnya tumpul (berlawanan arah) maka korelasinya negatif.
3. Posisi objek yang searah dengan suatu vektor peubah diinterpretasikan sebagai besarnya nilai peubah untuk objek yang searah dengannya. Semakin dekat letak objek dengan arah yang ditunjuk oleh suatu peubah maka semakin tinggi peubah tersebut untuk objek itu. Sedangkan jika arahnya berlawanan, maka nilainya rendah.
4. Kedekatan letak/posisi dua buah objek diinterpretasikan sebagai kemiripan sifat dua objek. Semakin dekat letak dua buah objek maka sifat yang ditunjukkan oleh nilai-nilai peubahnya semakin mirip.

Output dari teladan diatas adalah :

Singular values and variance accounted for

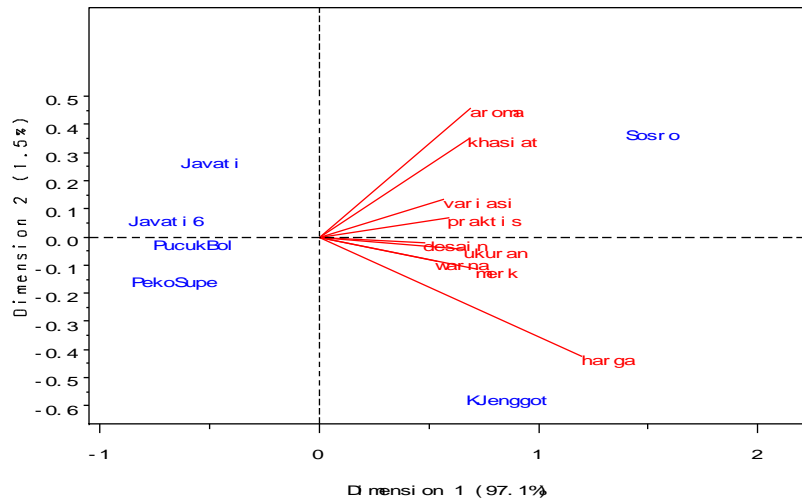
Singular Values	Percent	Cum %
4.5455	97.14	97.14
0.5607	1.48	98.61
0.4348	0.89	99.50
0.3047	0.44	99.94
0.1141	0.06	100.00
0.0000	0.00	100.00
0.0000	0.00	100.00
0.0000	0.00	100.00
0.0000	0.00	100.00

OBS / VARS ratio: 1.231934 Scale: 1

Biplot Factor Type

Symmetric

Biplot coordinates

	DIM1	DIM2
OBS Sosro	1.5247	0.3741
OBS Javati6	-0.6953	0.0702
OBS KJenggot	0.8706	-0.5639
OBS PekoSupe	-0.6600	-0.1447
OBS PucukBol	-0.5587	-0.0128
OBS Javati	-0.4812	0.2771
VAR ukuran	0.6655	-0.0419
VAR desain	0.4802	-0.0190
VAR harga	1.2012	-0.4256
VAR merk	0.7210	-0.1142
VAR warna	0.5364	-0.0854
VAR variasi	0.5713	0.1364
VAR khasiat	0.6828	0.3517
VAR aroma	0.6933	0.4584
VAR praktis	0.5903	0.0688

Keragaman yang diterangkan oleh sumbu utama 1 sebesar 97.14 dan sumbu utama 2 sebesar 1.48, sehingga secara keseluruhan keragaman yang dapat diterangkan oleh kedua sumbu tersebut sebesar 98.61%. Ratio skala yang digunakan dalam biplot adalah 1. Pada biplot terlihat ada empat pengelompokan merk yang cukup jelas. Sosro menempatkan diri sebagai merk dengan sifat-sifat yang lebih tinggi nilainya dari yang lain, seperti aroma, khasiat yang terkandung dan variasi rasanya. Cap kepala jenggot menyusul di belakangnya. Cap kepala jenggot dikenal sebagai merk teh dengan harga yang mahal. Keduanya dinilai mempunyai desain kemasan yang menarik.

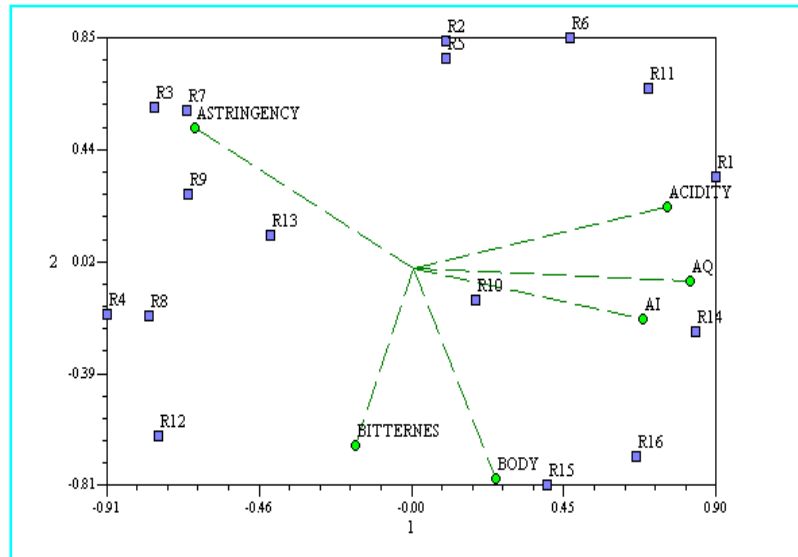
Kelompok berikutnya adalah merk pucuk bola dan peko super, yang digambarkan memiliki tingkat harga yang relatif murah dan warna seduhan yang tidak terlalu jelek, tapi dinilai memiliki desain kemasan yang paling buruk dari kelompok yang lain.

Kelompok terakhir adalah teh javati dan javati 6138. Kelompok ini dinilai sebagi merk-merk yang paling tidak populer, tidak dikenali. Namun demikian memiliki variasi rasa, yang menurut penilaian responden, cukup.

Jika dilihat perbandingan yang hanya melibatkan empat merk saja selain sosro dan cap kepala jenggot, javati dan cap pucuk bola mendapat penilaian yang lebih baik, terutama javati. Atribut menonjol dari merk pucuk bola adalah kepopuleran merk dan warna seduhan, meskipun untuk atribut warna masih lebih baik merk peko super.

Sedangkan javati terlihat menonjol hampir pada semua atribut. Bertolak belakang dengan kedudukan javati 6138 yang dinilai negatif pada semua atribut.

11.5. Latihan

1. Didapatkan hasil biplot seperti gambar dibawah ini

a. Apakahyang dimaksud dengan analisis biplot?
b. Sebutkan tahapan yang harus dilakukan dalam analisis biplot?
c. Perhatikan hasil analisis biplot diatas, yaitu mengenai 20 kode contoh produk parfum dengan peubah Aroma intensity, aroma quality, Body, Acidity, Bitternes, dan Astrigency. Interpretasilah hasil biplot tersebut dan buatkan kesimpulanya!

12

12. Analisis Korespondensi (*Correspondency Analysis*)

12.1. Pendahuluan

Analisis korespondensi adalah suatu ilmu yang mempelajari hubungan antara dua atau lebih peubah kualitatif, yaitu dengan teknik multivariate secara grafik yang digunakan untuk eksplorasi data dari sebuah tabel kontingensi. Analisis korespondensi ini meproyeksikan baris-baris dan kolom-kolom dari matriks data sebagai titk-titik ke dalam sebuah grafik berdimensi rendah dalam sebuah jarak Euclid. Analisis korespondensi seringkali digunakan untuk menetapkan kategori-kategori yang mirip dalam satu peubah, sehingga kategori-kategori tersebut dapat digabungkan menjadi satu kategori. Analisis ini juga bisa digunakan untuk menentukan kemungkinan hubungan antara dua gugus peubah.

Berdasarkan kegunaannya, analisis korespondensi dan analisis komponen utama memiliki kesamaan, yaitu suatu metode yang digunakan untuk mereduksi dimensi data menjadi dimensi yang lebih kecil dan sederhana. Sedangkan letak perbedaannya adalah bahwa analisis komponen utama lebih tepat untuk data dengan skala pengukuran kontinu sedangkan analisis korespondensi lebih tepat digunakan untuk data kategori.

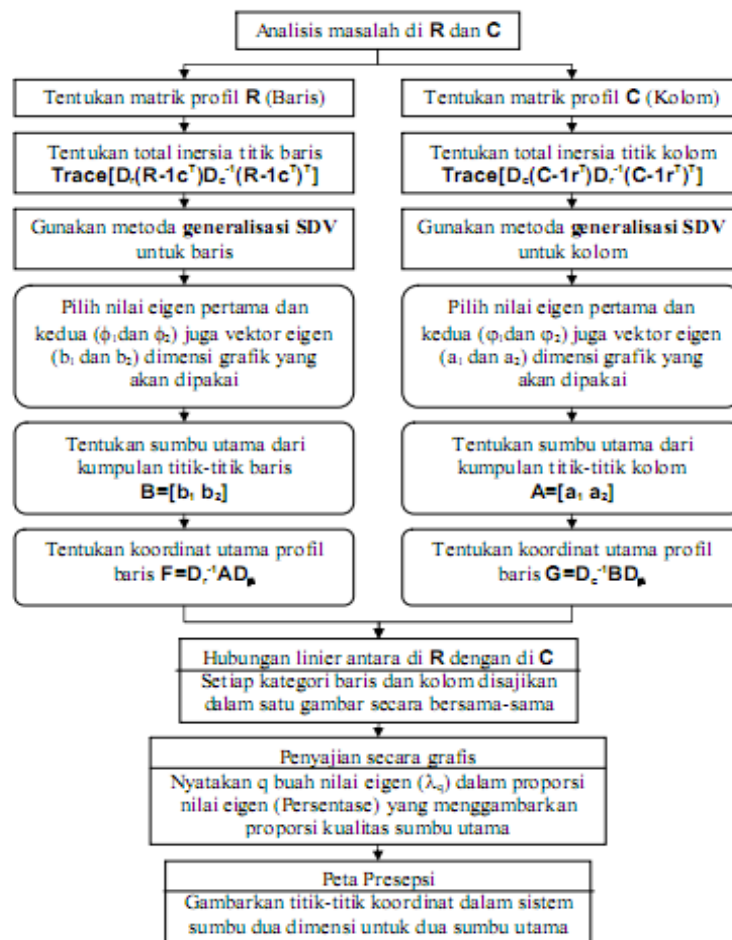
Dalam Analisis Korespondensi ada beberapa asumsi yang harus dipenuhi:

- ❑ Ukuran jarak Ki Kuadrat antar titik-titik (nilai kategori) analogi dengan konsep korelasi antar variabel.
- ❑ Variabel kolom yang tepat di variabel kategori baris diasumsikan homogen.
- ❑ Analisis Korespondensi adalah sebuah teknik nonparametrik yang tidak memerlukan pengujian asumsi seperti kenormalan, autokorelasi, multikolinearitas, heteroskedastisitas, linieritas sebelum melakukan analisis selanjutnya.
- ❑ Dimensi yang terbentuk dalam Analisis Korespondensi disebabkan dari kontribusi titik-titik dari dimensi yang terbentuk dan penamaan dari dimensinya subjektif dari kebijakan, pendapat dan error.
- ❑ Dalam Analisis Korespondensi variabel yang digunakan yaitu variabel diskrit (nominal/ordinal) yang mempunyai banyak kategori.

Beberapa kelebihan dan kekurangan analisis korespondensi yaitu :

- ❑ Kelebihan
    - ❑ Sangat tepat untuk menganalisis data variabel kategori ganda yang dapat digambarkan secara sederhana dalam data tabulasi silang.
    - ❑ Tidak hanya menggambarkan hubungan antar baris dengan kolom tetapi juga antar kategori dalam setiap baris dan kolom.
    - ❑ Memberikan tampilan grafik gab ungan dari kategori baris dan kolom dalam satu gambar yang berdimensi sama.
    - ❑ Cukup fleksibel untuk digunakan dalam data matrik berukuran besar.

- Kekurangan
  - Analisis ini tidak cocok untuk pengujian hipotesis tetapi sangat tepat untuk eksplorasi data.
  - Tidak mempunyai suatu metode khusus untuk menentukan atau memutuskan jumlah dimensi yang tepat.

Algoritma analisis korespondensi

12.2. Tabel Kontingensi Dua Arah

Jika X dan Y adalah dua peubah yang masing-masing mempunyai sebanyak a dan b kategori, maka dapat dibentuk suatu matriks data pengamatan $\mathbf{P}$ yang berukuran $a \times b$ dengan $p_{ij} \geq 0$ menyatakan frekuensi dari sel (i, j).

$$\mathbf{P} = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1b} \\ p_{21} & p_{22} & \cdots & p_{2b} \\ \vdots & \vdots & \ddots & \vdots \\ p_{a1} & p_{a2} & \cdots & p_{ab} \end{bmatrix}$$

Matriks $\mathbf{P}$ diatas juga dapat disajikan dalam bentuk table kontingensi, sebagai berikut :

Table 1. Table kontingensi dua arah

	Y_1	...	Y_j	...	Y_b	Total
X_1	n_{11}	...	n_{1j}	...	n_{1b}	$n_{1.}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
X_i	n_{i1}	...	n_{ij}	...	n_{ib}	$n_{i.}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
X_a	n_{a1}	...	n_{aj}	...	n_{ab}	$n_{a.}$
Total	$n_{.1}$	...	$n_{.j}$	...	$n_{.b}$	$n_{..}$

Keterangan :

$n_{i.} = \sum_{j=1}^{b} n_{ij}$; $i = 1,2, \dots, a$ peluang marginal X

$n_{.j} = \sum_{i=1}^{a} n_{ij}$; $j = 1,2, \dots, b$ Peluang marginal Y

$n_{..} = \sum_{i=1}^{a} \sum_{j=1}^{b} n_{ij}$ Total jumlah frekuensi dari matriks P

n_{ij} Frekuensi pengamatan ke-i baris pada kolom ke-j

Dari tabel kontingensi dua arah di atas dapat dibentuk matriks korespondensi sebagai berikut:

$$\mathbf{P}_{axb} = (p_{ij}) = (\frac{n_{ij}}{n})$$

Bila setiap elemen pada suatu baris dijumlahkan maka diperoleh vektor dari jumlah baris matriks P yaitu r' =P I = ($p_{1.}, \dots, p_{a.}$)'. Sehingga didapat Dr = diag (r) adalah diagonal matriks baris yaitu:

$$\mathbf{D}_r = diag(\mathbf{r}) = \begin{bmatrix} p_{1.} & 0 & \cdots & 0 \\ 0 & p_{2.} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & p_{a.} \end{bmatrix}$$

Dengan cara yang sama, akan didapat jumlah setiap kolom dari matriksnya menjadi vektor jumlah kolom dari matriks P yaitu $c = P' I = (p_{.1}, \dots, p_{.b})'$ sehingga didapat $D_c = diag(c)$ adalah diagonal matriks kolom sebagai berikut:

$$\mathbf{D}_c = diag(\mathbf{c}) = \begin{bmatrix} p_{.1} & 0 & \cdots & 0 \\ 0 & p_{.2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & p_{.b} \end{bmatrix}$$

12.3. Profil Baris dan Profil Kolom

Profil adalah proporsi dari setiap baris atau kolom Matriks Korespondensi yaitu setiap frekuensi pengamatan baris ke-i dan kolom ke-j dibagi dengan jumlah setiap total baris dan kolomnya masing-masing.

Matriks diagonal kolom dan baris diatas masing-masing berukuran $b\ x\ b$ dan $a\ x\ a$. Kemudian dapat dibentuk matriks R yang berukuran axb sebagai berikut:

$$\mathbf{R} = \mathbf{D}_r^{-1}\mathbf{P} = \begin{bmatrix} \frac{p_{11}}{p_{1.}} & \frac{p_{12}}{p_{1.}} & \cdots & \frac{p_{1b}}{p_{1.}} \\ \frac{p_{21}}{p_{2.}} & \frac{p_{22}}{p_{2.}} & \cdots & \frac{p_{2b}}{p_{2.}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{p_{a1}}{p_{a.}} & \frac{p_{a2}}{p_{a.}} & \cdots & \frac{p_{ab}}{p_{a.}} \end{bmatrix}$$

Matriks R disebut profil baris (row profile) dalam ruang berdimensi b , dengan jumlah unsur-unsur profil dari baris adalah sama dengan satu. Selanjutnya didefinisikan profil baris ke-i sebagai r_i yaitu:

$$\mathbf{r}_i = \left(\frac{p_{i1}}{p_{i.}}, \frac{p_{i2}}{p_{i.}}, \ldots, \frac{p_{ib}}{p_{i.}}\right)'$$

Sedangkan matriks C berukuran bxa adalah:

$$\mathbf{C} = \mathbf{D_c^{-1}P'} = \begin{bmatrix} \frac{p_{11}}{p_{.1}} & \frac{p_{12}}{p_{.1}} & \cdots & \frac{p_{a1}}{p_{.1}} \\ \frac{p_{21}}{p_{.2}} & \frac{p_{22}}{p_{.2}} & \cdots & \frac{p_{a2}}{p_{.2}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{p_{1b}}{p_{.b}} & \frac{p_{2b}}{p_{.b}} & \cdots & \frac{p_{ab}}{p_{.b}} \end{bmatrix}$$

Matriks C disebut sebagai profil kolom (column profile) dalam ruang berdimensi a, dimana jumlah unsur-unsur dari profil kolom sama dengan satu. Sehingga profil kolom ke- j sebagai c_j yaitu :

$$c_j = \left(\frac{p_{1j}}{p_{.j}}, \frac{p_{2j}}{p_{.j}}, \cdots, \frac{p_{aj}}{p_{.j}} \right)'$$

Untuk menampilkan profil-profil baris dan profil-profil kolom tersebut kedalam ruang dimensi Euclid yang berdimensi dua digunakan pendekatan jarak Ki Kuadrat, yaitu:

$$\chi^2 = \sum_{i=1}^{a} \sum_{j=1}^{b} \frac{\left(n_{ij} - \frac{n_{i.} n_{.j}}{n} \right)^2}{\frac{n_{i.} n_{.j}}{n}} = n \sum_{i=1}^{a} \sum_{j=1}^{b} \frac{(p_{ij} - p_{i.} p_{.j})^2}{p_{i.} p_{.j}}$$

$$= n \, \mathrm{tr}(\mathbf{E}) = n \sum_{i=1}^{m} \lambda_i^2$$

Misal diberikan suatu matriks korespondensi dengan D_r adalah matriks diagonal baris, D_c adalah matriks diagonal kolom, r merupakan vektor jumlah baris dan c adalah vektor jumlah kolom. Maka dapat dibentuk suatu matriks E sedemikian sehingga :

$$\mathbf{E} = \mathbf{D_r}^{-1} (\mathbf{P} - \mathbf{rc'}) \mathbf{D_c}^{-1} (\mathbf{P} - \mathbf{rc'})'$$

$\lambda_1^2 \geq \ldots \geq \lambda_m^2$ adalah nilai inersia atau akar ciri tak nol dari E dan m = rank (E) = rank $(P - rc') = (min(a, b) - 1)$ maka X^2 dapat juga dituliskan sebagai berikut:

$$X^2 = n \sum_{i=1}^{a} p_{i.} d_i^2$$

dengan $d_i^2 = (\mathbf{r}_i - \mathbf{c})' \mathbf{D}_c^{-1} (\mathbf{r}_i - \mathbf{c})$

12.4. Penguraian Nilai Singular

Untuk mereduksi dimensi data berdasarkan keragaman data (nilai eigen/inersia) terbesar dengan mempertahankan informasi optimum, diperlukan penguraian nilai singular. Penguraian nilai singular (*singular value decomposition*) merupakan salah satu konsep aljabar matriks dan konsep *eigen decomposition* yang terdiri dari nilai eigen λ dan vector eigen.

Teorema dekomposisi nilai singular :

Misalkan A matriks berukuran $m \times n$, maka ada matriks diagonal Σ berukuran $r \times r$ dimana $r \leq \min\{m, n\}$, matriks orthogonal U berukuran $m \times m$, matriks orthogonal V berukuran $m \times n$, sehingga $\boldsymbol{A} = \boldsymbol{U\Sigma V'}$.

Berdasarkan teorema dekomposisi nilai singular di atas, maka matriks yang akan di *singular value decomposition* matriks $\boldsymbol{U} = \boldsymbol{D_r}^{1/2}(\boldsymbol{P} - \boldsymbol{rc'})\boldsymbol{D_c}^{1/2}$ yang akan menghasilkan matriks $\boldsymbol{A}$ berukuran $a \times m$ dan matriks **B** berukuran $b \times m$, dan $\boldsymbol{\Lambda}$ merupakan suatu matriks yang elemen-elemennya adalah nilai singular, dimana nilai singular adalah akar dari nilai inersia.

12.4.1. Penguraian Nilai Singular Umum

Untuk menentukan anak ruang Euclid dan memproyeksikan semua profil baris ke dalam anak ruang Euclid digunakan penguraian nilai singular umum atau *Generalized Singular Value Decomposition* (GSVD)

Koordinat dari baris dan kolomnya ditentukan dengan menffunakan GSVD dari matriks $(\boldsymbol{P} - \boldsymbol{rc'})$, yaitu $\boldsymbol{A\Lambda B'}$, $\boldsymbol{\Lambda}$ merupakan matriks diagonal

yang mempunyai unsur-unsur diagonalnya nilai singular dari matriks $(\boldsymbol{P}-\boldsymbol{rc}')$, dimana merlaku $\boldsymbol{A}'\boldsymbol{D_r}^{-1}\boldsymbol{A}=\boldsymbol{I_m}$ dan $\boldsymbol{B}'\boldsymbol{D_c}^{-1}\boldsymbol{B}=\boldsymbol{I_m}$.

Tiap himpunan titik dapat dihubungkan dengan sumbu utama dari himpunan titik lainnya, yaitu :

	Rumusan koordinat baris	Rumusan Koordinat kolom
Analisis Profil Baris	$F=\boldsymbol{D_r}^{-1}\boldsymbol{A\Lambda}$	$G=\boldsymbol{D_c}^{-1}\boldsymbol{B}$
Analisis Profil Kolom	$F=\boldsymbol{D_r}^{-1}\boldsymbol{A}$	$G=\boldsymbol{D_c}^{-1}\boldsymbol{B\Lambda}$
Analisisn Baris dan Kolom	$F=\boldsymbol{D_r}^{-1}\boldsymbol{A\Lambda}$	$G=\boldsymbol{D_c}^{-1}\boldsymbol{B\Lambda}$

12.5. Nilai Inersia

Untuk mempresentasikan profil-profil baris dan profil-profil kolom ke dalam ruang dimensi $d\leq m$. koordinat dari i baris dari matriks yang dibentuk dengan mengambil d kolom pertama dari $\boldsymbol{F}=\boldsymbol{D_r}^{-1}\boldsymbol{A\Lambda}$, dan koordinat j profil kolom adalah j baris dari matriks yang dibentuk dengan mengambil k kolom pertama dari $\boldsymbol{G}=\boldsymbol{D_c}^{-1}\boldsymbol{B\Lambda}$. Karena total inersia mempresentasikan semua informasi dalam seluruh ruang adalah $tr(E)=n\sum_i^m\lambda_i$, maka pendekatan ruang berdimensi m dengan ruang berdimensi k dikatakan baik jika $\sum_i^d\lambda_i$ mendekati total inersia atau $\sum_i^m\lambda_i$ mendekati nol.

Nilai inersia menunjukan kontribusi dari baris ke-i pada inersia total. Inersia total adalah jumlah bobot kuadrat jarak titik-titik ke pusat, massa dan jarak yang didefenisikan.

Inersia total baris : $in\ (a) = \sum r_{i.}(r_{i.} - c)'D_c^{-1}\ (r_{i.} - c)$

Inersia total kolom: $in\ (b) = \sum c_{j.}(c_{j.} - r)'D_r^{-1}\ (c_{j.} - r)$

Jumlah bobot kuadrat koordinat titik-titik dalam sumbu utama ke-d pada tiap-tiap himpunan yaitu $\lambda_d^{\ 2}$ yang dinotasikan dengan λ_d . Nilai ini disebut sebagai inersia utama ke-d. Persamaan inersia utama baris dan kolom serta pusatnya dapat dinyatakan sebagai berikut :

Inersia utama baris adalah $\boldsymbol{F}^{'}\boldsymbol{D_r}\boldsymbol{F} = \boldsymbol{\Lambda}$

Bukti : $\boldsymbol{F}^{'}\boldsymbol{D_r}\boldsymbol{F} = \left(\boldsymbol{D_r}^{-1}\boldsymbol{A}\boldsymbol{\Lambda}\right)^{'}\boldsymbol{D_r}\left(\boldsymbol{D_r}^{-1}\boldsymbol{A}\right) = \boldsymbol{\Lambda}^{'}\boldsymbol{A}^{'}\left(\boldsymbol{D_r}^{-1}\right)^{-1}\boldsymbol{I}\boldsymbol{A} = \boldsymbol{\Lambda}'\boldsymbol{A}'\boldsymbol{D_r}^{-1}\boldsymbol{A}$

dengan menggunakan persamaan $\boldsymbol{A}'\boldsymbol{D_r}^{-1}\boldsymbol{A} = \boldsymbol{I_m}$, didapatkan $\boldsymbol{\Lambda}^{'}\boldsymbol{I_m} = \boldsymbol{\Lambda}'$. Karena matriks $\boldsymbol{\Lambda}'$ simetrik sehingga $\boldsymbol{\Lambda}^{'} = \boldsymbol{\Lambda}$, jadi $\boldsymbol{F}^{'}\boldsymbol{D_r}\boldsymbol{F} = \boldsymbol{\Lambda}$

Inersia utama baris adalah $\boldsymbol{G}^{'}\boldsymbol{D_c}\boldsymbol{G} = \boldsymbol{\Lambda}$

Bukti : $\boldsymbol{G}^{'}\boldsymbol{D_c}\boldsymbol{G} = \left(\boldsymbol{D_c}^{-1}\boldsymbol{B}\boldsymbol{\Lambda}\right)^{'}\boldsymbol{D_c}\left(\boldsymbol{D_c}^{-1}\boldsymbol{B}\right) = \boldsymbol{\Lambda}^{'}\boldsymbol{B}^{'}\left(\boldsymbol{D_c}^{-1}\right)^{-1}\boldsymbol{I}\boldsymbol{B} = \boldsymbol{\Lambda}'\boldsymbol{B}'\boldsymbol{D_c}^{-1}\boldsymbol{B}$

dengan menggunakan persamaan $\boldsymbol{B}'\boldsymbol{D_c}^{-1}\boldsymbol{B} = \boldsymbol{I_m}$, didapatkan $\boldsymbol{\Lambda}^{'}\boldsymbol{I_m} = \boldsymbol{\Lambda}'$. Karena matriks $\boldsymbol{\Lambda}'$ simetrik sehingga $\boldsymbol{\Lambda}^{'} = \boldsymbol{\Lambda}$, jadi $\boldsymbol{G}^{'}\boldsymbol{D_c}\boldsymbol{G} = \boldsymbol{\Lambda}$

Besaran $\lambda_1^{\ 2}, \ldots., \lambda_i^{\ 2}$ dapat diinterpretasikan sebagai besarnya kontribusi yang diberikan pada total inersia oleh masing-masing dimensi pertama, kedua, dan seterusnya, sehingga besaran relative untuk mengukur besarnya kehilangan informasi dapat dirumuskan sebagai berikut :

$$L = 1 - \frac{\sum_i^d \lambda_i^{\ 2}}{\sum_i^m \lambda_i^{\ 2}}$$

12.6. Koefisien Kontingensi

Untuk melihat keeratan hubungan atau kecendrungan antara variable satu dengan yang lainnya, digunakan rumusan kooefisien kontingensi sebagai berikut :

$$C = \sqrt{\frac{\chi^2}{N + \chi^2}}$$

dimana $\chi^2 =$ statistik uji Khi Kuadrat, dan N =banyaknya populasi sample,$0 \leq C \leq 1$

12.7. Ilustrasi

Pada suatu studi kasus, kepada 1500 responden (mahasiswa suatu perguruan tinggi) diminta untuk mengisi kusioner dengan peubah-peubah sebagai berikut :

1. *Code* : nomor urut responden
2. *Sex* : jenis kelamin responden
3. *Age* : umur peserta responden
4. *Family income* : pendapatan orang tua, dalam hal ini pendapatan dibagi menjadi 4
   kategori, yaitu 0-25000 ; 25001-50000 ; 50001-75000 ; 75000-inf
5. *GPA* : nilai indeks prestasi kumulatif yang diperoleh, dengan skala 1 sampai 4
6. *IQ* : nilai intelegensi
7. *Class standing* : dibagi dalam 4 kategori, yaitu mahasiswa baru (*freshman*), mahasiswa tingkat 2 (*sophomore*), mahasiswa tingkat 3 (*junior*), mahasiswa tingkat akhir (*senior*)
8. *Drink week* : banyaknya Alkohol yang dikonsumsi setiap minggunya (dalam satuan botol)
9. *Hardworker* : banyaknya kelas yang dihadiri, dengan rentangan 1-20
10. *Final grade* : nilai akhir yang diperoleh dari mata kuliah Psikologi

Dalam hal ini ingin dilihat apakah terdapat hubungan antara *family income* dengan *class standing.*

Hasil Analisis

Tabulasi silang antara peubah family income dan class standing disuatu perguruang tinggi dapat dilihat pada tabel *output* SPSS di bawah ini:

Correspondence Table

Family Income	Class Standing				
	Freshman	Sophomore	Junior	Senior	Active Margin
00000 - 25000	23	122	39	23	207
25001 - 50000	89	288	97	90	564
50001 - 75000	67	282	114	62	525
75001 - inf	34	104	36	30	204
Active Margin	213	796	286	205	1500

Dari tabel tabulasi silang di atas, dapat dilihat bahwa :

1. Responden terbanyak berasal dari *class standing* dengan kategori *sophomore* (mahasiswa tingkat 2)
2. Pada umumnya responden berasal dari keluarga dengan family income 25001-50000.

Row Profiles

Family Income	Class Standing				
	Freshman	Sophomore	Junior	Senior	Active Margin
00000 - 25000	.111	.589	.188	.111	1.000
25001 - 50000	.158	.511	.172	.160	1.000
50001 - 75000	.128	.537	.217	.118	1.000
75001 - inf	.167	.510	.176	.147	1.000
Mass	.142	.531	.191	.137	

Dari profil baris di atas dapat dilihat bahwa :

1. Nilai massa terbesar (0,531) terdapat pada class standing kategori sophomore yang merupakan modus pada data ini

2. Jika diperhatikan profil setiap kolom, maka dapat dilihat nilai terbesar untuk kolom:
   a. Freshman adalah pada baris family income dengan kategori ke 4 (75001-inf), hal ini menunjukkan bahwa secara umum mahasiswa dengan family income 75001-inf berasal dari mahasiswa baru (freshman).
   b. Sophomore adalah pada baris family income dengan kategori 1 (0-25000), yang menunjukkan bahwa umumnya mahasiswa tingkat 2 (sophomore) kebanyakan berasal dari keluarga dengan income paling rendah.
   c. Junior adalah pada baris familiy income dengan kategori 3 (50001-75000), yang menunjukkan bahwa mahasiswa tingkat 3 (junior) berasal dari keluarga dengan family income 50001-75000.
   d. Senior adalah pada baris familiy income dengan kategori 2 (25001-50000), yang menunjukkan bahwa mahasiswa tingkat akhir (senior) umumnya berasal dari keluarga dengan familiy income menengah yaitu 25001-50000.

Column Profiles

Family Income	Class Standing				
	Freshman	Sophomore	Junior	Senior	Mass
00000 - 25000	.108	.153	.136	.112	.138
25001 - 50000	.418	.362	.339	.439	.376
50001 - 75000	.315	.354	.399	.302	.350
75001 - inf	.160	.131	.126	.146	.136
Active Margin	1.000	1.000	1.000	1.000	

Dari profil kolom di atas dapat dilihat bahwa :

1. Nilai massa terbesar (0,376) terdapat pada family income dengan kategori income 25001-50000 yang merupakan modus pada data ini

2. Jika diperhatikan profil setiap baris, maka dapat dilihat nilai terbesar untuk baris :
   a. Family income kategori 1 adalah pada kolom class standing kategori sophomore, hal ini menunjukkan bahwa secara umum mahasiswa dengan family income terendah (0-25000) berasal dari mahasiswa tingkat 2 (sophomore)
   b. Family income kategori 2 adalah pada kolom class standing kategori senior, yang menunjukkan bahwa umumnya mahasiswa tingkat akhir (senior) kebanyakan berasal dari keluarga dengan income bekisar antara 25001 sampai 50000.
   c. Family income dengan kategori 3 adalah pada kolom class standing kategori Junior, yang menunjukkan bahwa keluarga dengan income berkisar antara 50001-75000 berasal mahasiswa tingkat 3 (junior).
   d. Family income dengan kategori 4 adalah pada kolom class standing kategori freshman, yang menunjukkan bahwa pada umumnya mahasiswa yang baru masuk berasal dari keluarga dengan income paling tinggi yaitu berkisar antara 75001-inf.

Jika kita perhatikan dari kedua profil di atas, baik profil baris maupun profil kolom, terlihat bahwa keduanya menyatakan satu hal yang sama meski dengan tampilan nilai-nilai yang berbeda.

Summary

Dimension	Singular Value	Inertia	Chi Square	Sig.	Proportion of Inertia		Confidence Singular Value	
					Accounted for	Cumulative	Standard Deviation	Correlation 2
1	.089	.008			.866	.866	.026	.011
2	.033	.001			.119	.985	.026	
3	.012	.000			.015	1.000		
Total		.009	13.825	.129[a]	1.000	1.000		

Pada tabel di atas terlihat bahwa *eigenvalue* (*inertia*) menunjukkan bahwa nilai varians yang dapat dijelaskan adalah sebesar 0.009 atau 0.9% saja. Jika melihat nilai *Chi-Square*, dimana nilai p_value sebesar 0,129 (> 0,01) hal ini memperlihatkan bahwa tidak terdapat hubungan antara family income dengan class standing..

Korelasi kanonik maksimum (yang merupakan interpretasi dari *Singular Value* yang merupakan akar kuadrat dari *Eigenvalue*) antar kategori dari variabel-variabel dalam analisis untuk setiap dimensi adalah 0.089 untuk dimensi pertama (terbesar), 0.033 untuk dimensi kedua (juga merupakan yang kedua terbesar), dan 0.012 untuk dimensi ketiga, sedangkan dari *eigenvector* yang diperoleh dalam analisis (*proportion of inertia*), maka dengan 3 faktor dapat dinyatakan bahwa keragaman yang dapat diterangkan adalah sebesar 100% dengan rincian sebagai berikut:

1. Faktor pertama dengan *eigenvalue* sebesar 0.008 mampu menerangkan keragaman data sebesar 86.6%.
2. Faktor kedua dengan *eigenvalue* sebesar 0.001 mampu menerangkan keragaman data sebesar 11.9% (total dengan faktor pertama adalah 98.5 %)
3. Faktor ketiga dengan *eigenvalue* sebesar 0.000 mampu menerangkan keragaman data sebesar 1.5% (total dengan sebelumnya menjadi 100%).

Maka dari rincian di atas, cukup menggunakan 2 dimensi saja karena 2 dimensi ini sudah sangat baik dalam menerangkan keragaman data, yakni sebesar 98.5%.

Confidence Row Points dan *Confidence Column Points* di bawah ini memperlihatkan standard deviasi dari skor-skor baris dan kolom (nilai-nilai ini digunakan sebagai koordinat untuk memplot *Correspondence Map* dan juga digunakan untuk menilai presisinya).

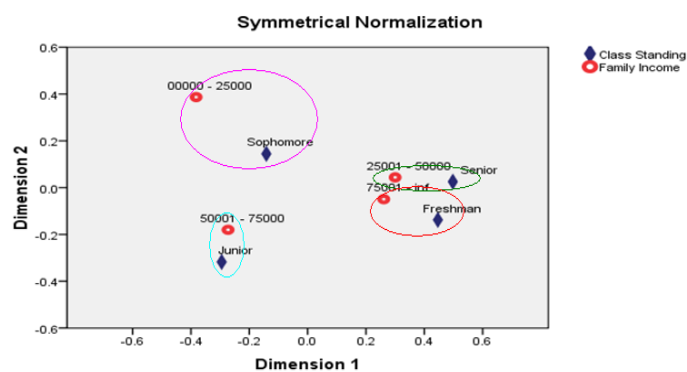
Confidence Row Points

Family Income	Standard Deviation in Dimension		Correlation
	1	2	1-2
00000 - 25000	.223	.183	.442
25001 - 50000	.087	.146	-.165
50001 - 75000	.112	.097	-.513
75001 - inf	.221	.401	-.014

Confidence Column Points

Class Standing	Standard Deviation in Dimension		Correlation
	1	2	1-2
Freshman	.189	.310	.171
Sophomore	.088	.072	.419
Junior	.195	.163	-.354
Senior	.181	.322	.006

Pada dua tabel di atas dapat dilihat bahwa standard deviasi yang besar menunjukkan presisi rendah rendah bagi suatu poin atau kategori untuk ditempatkan pada dimensi atau faktor tertentu. Sebaliknya semakin rendah nilai standard deviasinya, maka semakin baik presisinya.

Jika kita lihat presisi tersebut pada ke dua tabel di atas, maka family income untuk kategori 1 dan 4 berada pada dimensi yang relatif sama (berdekatan), begitu pula dengan class standing untuk kategori freshman dan senior.

Dari figur di atas dapat dilihat bahwa pada umumnya mahasiswa tingkat 2 berasal dari keluarga dengan income antara 0-25000, mahasiswa tingkat 3 berasal dari keluarga dengan income berkisar antara 50001-75000, mahasiswa tingkat akhir berasal dari keluarga dengan pendapatan 25001-50000, sedangkan mahasiswa baru kebanyakan berasal dari keluarga dengan pendapatan tertinggi yakni lebih dari 75000. Hal ini juga sesuai dengan profil baris dan profil kolom yang telah dipaparkan di bagian awal analisis korespondensi.

12.8. Aplikasi SAS

Dengan menggunakan analisis korespondensi, bisa didapatkan plot dua dimensi yang bisa menjelaskan asosiasi antar kategori setiap peubah dengan cepat, dibandingkan dengan melihat kembali angka-angka yang ada di tabel kontingensi. Untuk melakukan analisis korespondensi pada SAS menggunakan PROC CORRESP. Penggunaan prosedur tersebut bisa melibatkan dua struktur data yang berbeda. Struktur data pertama adalah kolom-kolom masih berupa varibel kategorik. Artinya data yang dimiliki masih berupa data mentah. Struktur data yang lain, adalah jika kita sudah memiliki rangkuman berupa tabel kontingensi.

Misalkan saja kita memiliki data 140 transaksi jual beli valuta asing yang dilakukan oleh perusahaan domestik dengan bank domestik. Ada sebanyak 4 perusahaan yang biasa melakukan penjualan valuta asing hasil ekspor produk mereka ke negara-negara tetangga, yaitu perusahaan A, B, C, dan D. Serta ada 4 bank utama yang menjadi tempat perusahaan-perusahaan tersebut menjual, yaitu W, X, Y, dan Z. dalam hal ini kita memiliki dua peubah kategorik, yaitu nama perusahaan penjual valuta asing dan nama bank tempat menjual valuta asing

Data yang diperoleh dari 140 transaksi terakhir adalah sebagai berikut (nomor yang ada merupakan nomor urut transaksi)

1	B	W	21	C	W	41	C	W	61	D	X	81	C	W	101	A	W	121	D	W
2	C	Y	22	B	X	42	B	X	62	D	X	82	A	Z	102	C	X	122	D	Y
3	B	X	23	B	Z	43	C	W	63	A	Z	83	A	Y	103	B	W	123	C	W
4	C	W	24	B	X	44	B	W	64	C	Y	84	A	X	104	B	Z	124	B	Y
5	B	X	25	D	X	45	C	Z	65	D	Y	85	D	W	105	A	Z	125	A	Y
6	D	Z	26	B	W	46	B	W	66	B	Y	86	A	X	106	B	Z	126	B	W
7	A	Z	27	B	X	47	A	Y	67	B	X	87	B	W	107	D	Z	127	C	Y
8	A	Y	28	A	X	48	D	X	68	B	Y	88	A	X	108	A	Y	128	C	W
9	D	Y	29	B	X	49	A	X	69	A	Y	89	C	X	109	C	X	129	C	Z
10	B	Y	30	A	Y	50	B	Z	70	C	W	90	C	X	110	D	X	130	D	Z
11	D	X	31	A	X	51	B	Z	71	B	W	91	C	Z	111	C	W	131	B	X
12	A	X	32	C	Y	52	D	X	72	C	Y	92	C	X	112	C	X	132	B	W
13	C	Y	33	D	Y	53	B	Y	73	D	Z	93	D	X	113	D	Z	133	A	W
14	A	Y	34	C	X	54	A	W	74	B	W	94	B	X	114	B	W	134	B	W
15	A	X	35	C	Y	55	C	X	75	D	Y	95	B	X	115	D	X	135	A	X
16	C	W	36	C	W	56	C	X	76	C	W	96	D	X	116	C	Y	136	B	X
17	B	W	37	B	Y	57	C	X	77	D	Y	97	C	Z	117	C	W	137	B	W
18	B	Y	38	C	Y	58	C	W	78	B	Y	98	D	W	118	A	Y	138	D	Y
19	D	X	39	C	X	59	D	Y	79	A	X	99	C	Z	119	D	X	139	B	W
20	B	W	40	B	W	60	B	W	80	A	Z	100	C	W	120	A	X	140	C	X

Perintah menyiapkan dan melakukan analisis korespondensi yang diberikan adalah sebagai berikut:

```
data contoh;
input no Penjual$ Pembeli$ @@;
cards;
1   B   W       21      C       W       41      C       W       61      D
    X   81      C       W       101     A       W       121     D       W
2   C   Y       22      B       X       42      B       X       62      D
    X   82      A       Z       102     C       X       122     D       Y
3   B   X       23      B       Z       43      C       W       6       A
    Z   83      A       Y       103     B       W       123     C       W
4   C   W       24      B       X       44      B       W       64      C
    Y   84      A       X       104     B       Z       124     B       Y
5   B   X       25      D       X       45      C       Z       65      D
    Y   85      D       W       105     A       Z       125     A       Y
6   D   Z       26      B       W       46      B       W       66      B
    Y   86      A       X       106     B       Z       126     B       W
7   A   Z       27      B       X       47      A       Y       67      B
    X   87      B       W       107     D       Z       127     C       Y
8   A   Y       28      A       X       48      D       X       68      B
    Y   88      A       X       108     A       Y       128     C       W
9   D   Y       29      B       X       49      A       X       69      A
    Y   89      C       X       109     C       X       129     C       Z
10  B   Y       30      A       Y       50      B       Z       70      C
    W   90      C       X       110     D       X       130     D       Z
11  D   X       31      A       X       51      B       Z       71      B
    W   91      C       Z       111     C       W       131     B       X
12  A   X       32      C       Y       52      D       X       72      C
    Y   92      C       X       112     C       X       132     B       W
13  C   Y       33      D       Y       53      B       Y       73      D
    Z   93      D       X       113     D       Z       133     A       W
14  A   Y       34      C       X       54      A       W       74      B
    W   94      B       X       114     B       W       134     B       W
15  A   X       35      C       Y       55      C       X       75      D
    Y   95      B       X       115     D       X       135     A       X
16  C   W       36      C       W       56      C       X       76      C
    W   96      D       X       116     C       Y       136     B       X
17  B   W       37      B       Y       57      C       X       77      D
    Y   97      C       Z       117     C       W       137     B       W
18  B   Y       38      C       Y       58      C       W       78      B
    Y   98      D       W       118     A       Y       138     D       Y
19  D   X       39      C       X       59      D       Y       79      A
    X   99      C       Z       119     D       X       139     B       W
20  B   W       40      B       W       60      B       W       80      A
    Z   100     C       W       120     A       X       140     C       X

;
proc corresp outc=hasil;
tables penjual, pembeli;

%plotit (data=hasil, datatype=corresp, color='BLUE');
run;
```

Dengan perintah di atas diperoleh output sebagai berikut.

```
                         The CORRESP Procedure
                  Inertia and Chi-Square Decomposition

   Singular    Principal     Chi-              Cumulative
     Value      Inertia     Square   Percent     Percent    20   40   60   80  100
                                                            ----+----+----+----+---
   0.31572    0.09968    13.9548     98.90       98.90     ***********************
   0.03277    0.00107     0.1503      1.07       99.97
   0.00552    0.00003     0.0043      0.03      100.00

     Total    0.10078    14.1095    100.00

   Degrees of Freedom = 9
```

Hasil di atas menggambarkan berapa persen kemampuan plot korespondensi memberikan informasi dari data awal. Dari data contoh,

diperoleh bahwa dengan menggunakan 2 dimensi pertama, mampu menerangkan data asal sebesar 99.97%. Ini berarti bahwa dalam kasus ini, informasi yang diberikan hampir sempurna.

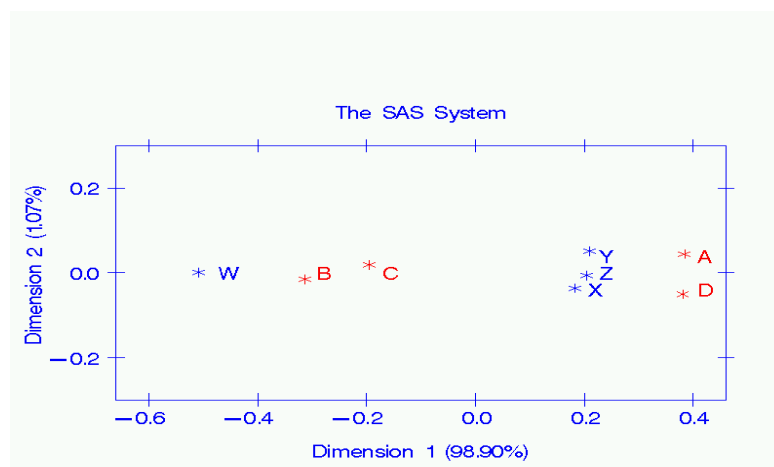
Berikutnya adalah koordinat dari setiap kategori peubah

Row Coordinates

	Dim1	Dim2
A	0.3841	0.0449
B	-0.3129	-0.0148
C	-0.1943	0.0189
D	0.3810	-0.0499

Column Coordinates

	Dim1	Dim2
W	-0.5078	0.0017
X	0.1826	-0.0357
Y	0.2099	0.0513
Z	0.2042	-0.0064

Koordinat di ataslah yang nantinya akan dijadikan sebagai titik-titik pada plot dua dimensi hasil analisis korespondensi. Plot yang diperoleh dengan koordinat di atas adalah:

Berdasarkan data di atas terlihat bahwa perusahaan A dan D lebih sering menjual valuta asingnya ke bank X, Y atau Z, daripada menjualnya ke bank W.

Output lain berisi inersia dari setiap titik terhadap pembentukan kedua dimensi.

```
                Summary Statistics for the Row Points

                         Quality       Mass    Inertia

               A          0.9999     0.2000     0.2968
               B          0.9997     0.3071     0.2992
               C          0.9986     0.2929     0.1109
               D          1.0000     0.2000     0.2930


         Partial Contributions to Inertia for the Row Points

                                  Dim1       Dim2

                     A          0.2960     0.3762
                     B          0.3018     0.0628
                     C          0.1109     0.0976
                     D          0.2913     0.4634


               Summary Statistics for the Column Points

                         Quality       Mass    Inertia

               W          1.0000     0.2786     0.7127
               X          0.9996     0.3357     0.1154
               Y          0.9999     0.2429     0.1125
               Z          0.9960     0.1429     0.0594

       Partial Contributions to Inertia for the Column Points

                                  Dim1       Dim2

                     W          0.7206     0.0007
                     X          0.1124     0.3993
                     Y          0.1073     0.5945
                     Z          0.0598     0.0055
```

Selain menggunakan data mentah, data dalam bentuk tabel kontingensi juga dapat dianalisis. Untuk mendapatkan tabel kontingensi dari data mentah di atas, bisa dilakukan menggunakan prosedur PROC FREQ pada SAS/BASE.

Perintah yang diberikan adalah sebagai berikut

```
proc freq data=contoh;
tables penjual*pembeli/norow nocol nopercent;
run;
```

Hasilnya adalah sebagai berikut

```
                              The FREQ Procedure

                         Table of Penjual by Pembeli

               Penjual      Pembeli

               Frequency|W        |X        |Y        |Z        |  Total
                      ---+--------+--------+--------+--------+
               A        |       3 |      11 |       9 |       5 |     28
                      --+--------+--------+--------+--------+
               B        |      18 |      12 |       8 |       5 |     43
                      --+--------+--------+--------+--------+
               C        |      15 |      12 |       9 |       5 |     41
                      --+--------+--------+--------+--------+
               D        |       3 |      12 |       8 |       5 |     28
                      --+--------+--------+--------+--------+
               Total          39       47       34       20      140
```

Selanjutnya, seandainya data yang mentah tidak kita miliki, namun kita memiliki data yang sudah berupa tabel kontingensi di atas, maka perintah yang diberikan adalah

```
data contoh2;
input penjual$ pembeli$ frekuensi;
cards;
A     W        3
A     X        11
A     Y        9
A     Z        5
B     W        18
B     X        12
B     Y        8
B     Z        5
C     W        15
C     X        12
C     Y        9
C     Z        5
D     W        3
D     X        12
D     Y        8
D     Z        5
;
proc corresp data=contoh2 outc=hasil2;
tables penjual, pembeli;
weight frekuensi;
run;

%plotit (data=hasil2, datatype=corresp);
run;
```

Hasil yang diperoleh sama persis dengan hasil pada kasus menggunakan data mentah.

13

13.Korelasi Kanonik (*Canonical Correlation*)

13.1. Pendahuluan

Dalam banyak penelitian seringkali ditemukan respon terdiri dari banyak (lebih dari satu) peubah, sehingga mengharuskan peneliti untuk menggunakan metode analisis yang sesuai. Dalam kasus penelitian yang melibatkan banyak/multi respon, analisis statistika yang sesuai atau cocok digunakan adalah analisis peubah ganda (*multivariate*).

Analisis peubah ganda (*multivariate*) merupakan salah satu teknik statistika yang digunakan untuk menganalisis data peubah ganda. Data peubah ganda yang dimaksud adalah data yang terdiri dari lebih dari satu peubah/peubah bebas (*independent*) dan lebih dari satu peubah/peubah tak bebas (*dependent*). Jenis data yang digunakan dalam analisis peubah ganda dapat terdiri dari data non metrik (skala nominal dan ordinal) dan data metrik (skala interval dan rasio).

Salah satu teknik analisis peubah ganda yang sering digunakan ketika seorang peneliti ingin menguji hubungan (korelasi) antara beberapa peubah independen dengan beberapa peubah dependen adalan analisis korelasi kanonik. Beberapa peubah independen disebut gugus peubah independen sedangkan beberapa peubah dependen yang disebut gugus peubah dependen. Konsep korelasi dalam analisis peubah ganda merupakan perluasan dari konsep korelasi pada

univariate yang responnya merupakan peubah tunggal. Hal yang membedakan diantara keduanya adalah jumlah dari peubah respon. Dalam *univariate*, peubah respon hanya terdiri dari satu peubah sedangkan dalam *multivariate* memiliki lebih dari satu peubah respon.

Jenis korelasi yang dikenal dalam kasus *univariate* adalah korelasi sederhana, korelasi parsial dan korelasi berganda. Dalam kasus *multivariate*, analisis korelasi lebih dikenal dengan istilah analisis korelasi kanonik. Analisis korelasi kanonik tidak sesederhana korelasi sederhana, parsial atau berganda. Hal ini karena dalam analisis korelasi kanonik yang dicari adalah korelasi antar gugus peubah independen dan gugus peubah dependen bukan korelasi antar peubah independen dan dependen. Berikut ini adalah contoh penelitian yang menggunakan korelasi kanonik dalam analisisnya.

1. Seorang dokter ingin mengetahui adakah hubungan antara gaya hidup dan kebiasaan makan dengan kesehatan pasienyang diukur dengan peubah hipertensi, berat badan, tingkat ketegangan dan *anxiety*.
2. Manajer pemasaran suatu perusahaan ingin mengetahui apakah ada hubungan antara jenis produk yang dibeli dan gaya hidup konsumen dan kepribadian konsumen.
3. Seorang direktur bank ingin mengetahui hubungan antara penggunaan kartu kredit (diukur dengan jumlah kartu kredit yang dimiliki dan rata-rata uang yang dibelanjakan lewat kartu kredit perbulan) dengan karakteristik konsumen (diukur dengan besarnyajumlah keluarga dan *ncome* keluarga).

Dalam menggunakan analisis korelasi kanonik, terdapat beberapa tahapan analisis yang umum dilakukan, seperti:

1. Melakukan uji data dan uji asumsi analisis korelasi kanonik
2. Menentukan fungsi kanonik dan ukuran kesesuaiannya

3. Menginterpretasikan hasil analisis korelasi kanonik
4. Melakukan validasi hasil analisis korelasi kanonik.

13.2. Analisis Korelasi Kanonik

Analisis korelasi kanonik adalah salah satu teknik analisis statistika yang digunakan untuk melihat hubungan antara segugus peubah dependen (Y_1, Y_2,....., Y_p) dengan segugus peubah independen (X_1, X_2,....., X_q).

Fokus perhatian dalam analisis korelasi kanonik adalah korelasi (hubungan), sehingga pada dasarnya kedua himpunan tidak perlu dibedakan menjadi kelompok vaiabel *independent* dan *dependent*. Pemberian label X dan Y hanya untuk membedakan kedua himpunan peubah tersebut.

Analisis korelasi kanonik berfokus pada korelasi antara kombinasi linear dari gugus peubah dependen dengan kombinasi linear dari gugus peubah independen. Ide utama dari analisis ini adalah mencari pasangan dari kombinasi linear ini yang memiliki korelasi terbesar. Pasangan dari kombinasi linear ini disebut fungsi/peubah kanonik dan korelasinya disebut korelasi kanonik.

Tujuan dari analisis korelasi kanonik adalah :

1. Mengukur tingkat keeratan hubungan antara segugus peubah dependen dengan segugus peubah independen
2. Menguraikan struktur hubungan di dalam gugus peubah dependen maupun dalam gugus peubah independen

13.2.1. Uji Data dan Uji Asumsi

Asumsi-asumsi yang harus dipenuhi dalam analisis korelasi kanonik adalah (Nugroho, 2008) :

1. Linieritas, yaitu keadaan dimana hubungan antar peubah dependen dengan peubah independen bersifat linier.

   Sebagai contoh : jika ada peubah promosi dan penjualan, maka seharusnya korelasi antara kedua peubah bersifat linier. Dalam arti makin besar pengeluaran untuk promosi, maka makin tinggi juga penjualannya.
2. Perlunya *multivariate normality* untuk menguji signifikansi setiap fungsi kanonik. Karena pengujian normalitas secara multivariate sulit dilakukan, maka cukup dilakukan uji normalitas untuk setiap peubah. Asumsi yang digunakan adalah jika secara individu sebuah peubah memenuhi kriteria normalitas, maka secara keseluruhan juga akan memenuhi asumsi normalitas.
3. Tidak ada multikolinearitas antar anggota kelompok peubah, baik peubah dependen maupun peubah independen. Sebagai contoh, jika peubah dependen terdiri dari penjualan dan biaya produksi, maka seharusnya tidak ada korelasi yang kuat dan nyata antara peubah penjualan dengan peubah biaya produksi. Jika ada korelasi, dinamakan terdapat multikolinieritas. Jika angka korelasi tersebut besar maka dapat dilakukan pengurangan salah satu peubah, misal salah satu dari peubah penjualan atau biaya produksi dapat dihilangkan.

Adanya beberapa asumsi yang harus dipenuhi pada analisis korelasi kanonik, sebelum melakukan analisis perlu dilakukan pengujian untuk data yang akan dianalisis dalam dua tahap, yaitu: (1) Uji data dan (2) Uji asumsi. Uji data untuk analisis multivariat meliputi: uji data yang tidak

lengkap (*missing values*) dan uji data pencilan (*outlier*). Uji asumsi meliputi uji normalitas, homoskedastisitas dan linieritas.

Untuk mendeteksi data pencilan dalam analisis *mulitivariate* dilakukan dengan menggunakan jarak Mahalanobis (*Mahalanobis* D^2). Jarak Mahalanobis adalah ukuran yang menyatakan jarak nilai setiap kasus dari rata-rata seluruh kasus. Jarak Mahalanobis yang besar menandakan nilai ekstrem suatu kasus terhadap satu atau lebih peubah.

Uji asumsi linieritas dilakukan dengan melakukan analisis korelasi-regresi linier antara peubah independen dan peubah dependen. Uji kenormalan dilakukan dengan menggunakan plot distribusi normal dan uji Kolmogorof-Smirnov dengan Hipotesis nol (H_0) semua peubah berdistribusi normal dan hipotesis tandingannya (H_1) peubah tidak normal. Untuk uji homokedasitas dilakukan dengan uji Levene.

Setelah dilakukan uji asumsi, sebaiknya semua asumsi terpenuhi. Dengan kata lain, jika terdapat peubah yang tidak memenuhi asumsi maka sebaiknya peubah tersebut dihilangkan/dibuang. Namun hal tersebut tidak dengan mudah dapat dilakukan karena pembuangan peubah akan menyebabkan tujuan analisis tidak dapat dipenuhi serta menyebabkan sifat komposit peubah tidak dapat dipertahankan. Transformasi data untuk memperbaiki data, juga harus mempertimbangkan makna satuan hasil transformasi. Misalkan jika diputuskan untuk melakukan transformasi dengan merubah data menjadi log-natural, maka interpretasi hasil tidak lagi melibatkan satuan unit asal tetapi sudah melibatkan satuan baru yang dalam banyak kasus dapat mengakibatkan kesimpulan yang berbeda.

Perbaikan data pada analisis *multivariate* tidak semudah pada analisis *univariate*. Seringkali perbaikan data diperlukan untuk peubah yang

satu tetapi tidak untuk peubah yang lain. Selain itu perbaikan untuk memenuhi asumsi yang satu sering kali menyebabkan asumsi lain dilanggar. Praktisnya, tujuan dari dilakukannya uji asumsi adalah agar peneliti lebih hati-hati ketika melakukan interpretasi hasil analisis daripada sebagai alat untuk menghilangkan peubah atau melakukan transformasi data.

13.2.2. Penentuan Fungsi Kanonik dan Pendugaan Koefisien Kanonik

Penentuan fungsi kanonik bisa dilakukan dengan menggunakan matriks *covarian* atau matriks korelasi. Hal yang membedakan keduanya adalah data yang digunakan dalam analisis. Matriks korelasi digunakan jika data sudah dibakukan (memiliki satuan yang sama), sedangkan matriks *covarian* menggunakan data sebenarnya (data tidak dibakukan dan memiliki satuan yang sama). Proses penentuan fungsi kanonik dari kedua jenis matriks tersebut sama.

1. Penentuan fungsi kanonik dari fungsi covarian

Misalkan ingin dibuat hubungan antara gugus peubah dependen Y_1, Y_2, …, Y_p yang dinotasikan dengan vektor peubah acak **Y**, dengan gugus peubah independen X_1, X_2, …,X_q yang dinotasikan dengan vektor peubah acak **X**, dimana $p \leq q$.

Misalkan karakteristik dari vektor peubah acak **X** dan **Y** adalah sebagai berikut :

$E(X) = \mu_x$ $\quad$ $Cov(Y) = \sum_{yy}$

$E(Y) = \mu_y$ $\quad$ $Cov(X) = \sum_{xx}$

$Cov(X,Y) = \sum_{xy} = (\sum_{yx})^t$

Kombinasi linear dari kedua gugus peubah dapat dituliskan sebagai berikut :

$W = \underline{a}^t X = a_1X_1 + a_2X_2 + ...+a_pX_q$

$V = \underline{b}^t Y = b_1Y_1 + b_2Y_2 + ...+b_qY_p$

$Var\ (W) = \underline{a}^t Cov(X)\underline{a} = \underline{a}^t \sum_{xx} \underline{a}$

$Var\ (V) = \underline{b}^t Cov(Y)\underline{b} = \underline{b}^t \sum_{yy} \underline{b}$

$Cov\ (W,V) = \underline{a}^t Cov(X,Y)\underline{b} = \underline{a}^t \sum_{xy} \underline{b}$

Dengan menggunakan **ketaksamaan Cauchy-Schwarz atau metode langrange,** vektor koefisien $\underline{a}$ dan $\underline{b}$ dapat diperoleh dengan cara mencari $\rho_1^2 > \rho_2^2 >..... > \rho_k^2$ yang merupakan Akar ciri dari matriks $\sum_{yy}^{-1/2} \sum_{yx} \sum_{xx}^{-1} \sum_{xy} \sum_{yy}^{-1/2}$ yang berpadanan dengan vektor ciri $\underline{f}_1, \underline{f}_2, \ldots, \underline{f}_k$. $\rho_1^2 > \rho_2^2 >..... > \rho_k^2$ juga merupakan akar ciri dari matriks $\sum_{xx}^{-1/2} \sum_{xy} \sum_{yy}^{-1} \sum_{yx} \sum_{xx}^{-1/2}$ yang berpadanan dengan vektor ciri $\underline{e}_1, \underline{e}_2, \ldots, \underline{e}_k$ sehingga vektor koefisien $\underline{a}$ dan $\underline{b}$ dapat diperoleh sebagai berikut :

$$a_1 = e_1\Sigma_{11}^{-1/2} \qquad b_1 = f_1\Sigma_{22}^{-1/2}$$

$$a_2 = e_2\Sigma_{11}^{-1/2} \qquad b_2 = f_2\Sigma_{22}^{-1/2}$$

$$\ldots \qquad \ldots$$

$$a_k = e_k\Sigma_{11}^{-1/2} \qquad b_k = f_k\Sigma_{22}^{-1/2}$$

Korelasi kanonik diperoleh dengan menghitung :

$$Corr(W,V) = \frac{\underline{a}'\sum_{X,Y}\underline{b}}{\sqrt{\underline{a}'\sum_{X.X}\underline{a}}\sqrt{\underline{b}'\sum_{Y.Y}\underline{b}}}$$

Didefinisikan pasangan pertama dari peubah kanonik adalah kombinasi linear W1, V1 yang memiliki ragam satu dan korelasi terbesar; pasangan kedua dari peubah kanonik adalah kombinasi linear W2, V2 yang memiliki ragam satu dan korelasi terbesar kedua serta tidak berkorelasi dengan peubah kanonik yang pertama dan pasangan ke-k dari peubah kanonik adalah kombinasi linear Wk, Vk yang memiliki ragam satu dan korelasinya terbesar ke-k serta tidak berkorelasi dengan peubah kanonik 1, 2, …, k-1.

Dengan demikian dapat dituliskan sebagai berikut :

- Fungsi Kanonik Pertama :

  $W_1 = \underline{a}_1^{t}X$ Var $(W_1) = 1$

  $V_1 = \underline{b}_1^{t}Y$ Var $(V_1) = 1$

  maksimum Corr $(W_1,V_1) = \rho_1$

- Fungsi Kanonik Kedua :

  $W_2 = \underline{a}_2^{t}X$ Var $(W_2) = 1$ Cov $(W_1,W_2) = 0$

  $V_2 = \underline{b}_2^{t}Y$ Var $(V_2) = 1$ Cov $(V_1,V_2) = 0$

  Cov (W_1,V_2) =Cov $(W_2,V_1) = 0$

  maksimum Corr $(W_2,V_2) = \rho_2$

- Fungsi Kanonik ke-k

  $W_k = \underline{a}_k^{t}X$ Var $(W_k) = 1$ Cov $(W_1,W_k) = 0$ $k \neq 1$

  $V_k = \underline{b}_k^{t}Y$ Var $(V_k) = 1$ Cov $(V_1,V_k) = 0$ $k \neq 1$

  Cov (W_1,V_k) =Cov $(W_k,V_1) = 0$ $k \neq 1$

  maksimum Corr $(W_k,V_k) = \rho_k$

2. Penentuan fungsi kanonik dari Fungsi Korelasi

Matriks Korelasi digunakan jika peubah yang akan dianalisis sudah dibakukan atau dengan kata lain sudah memiliki satuan yang sama. Berikut adalah matriks korelasi dari gugus peubah independen dan gugus peubah dependen.

$$R = \begin{pmatrix} R_{yy} & R_{yx} \\ R_{xy} & R_{xx} \end{pmatrix}$$

Dimana :

R_{yy} adalah matriks korelasi sampel y (pxp)

R_{yx} adalah matriks korelasi sampel antara y dan x (pxq)

R_{xx} adalah matriks korelasi sampel x (qxq)

Vektor Koefisien c dan d diperoleh sbb :

dari persamaan

$| R_{yy}^{-1}R_{yx}R_{xx}^{-1}R_{xy} - r^2I | = 0$ dan
$| R_{xx}^{-1}R_{xy}R_{yy}^{-1}R_{yx} - r^2I | = 0$

diperoleh akar ciri yang sama yaitu $r_1^2, r_2^2, \ldots, r_s^2$ tetapi vektor ciri yang berbeda yaitu

$(R_{yy}^{-1}R_{yx}R_{xx}^{-1}R_{xy} - r^2I)c = 0$
$(R_{xx}^{-1}R_{xy}R_{yy}^{-1}R_{yx} - r^2I|)d = 0$

Hubungan antara vektor ciri c dan d dengan vektor ciri e dan f yang dihasilkan matriks varian kovarian adalah:

$c = D_y f$ dan $d = D_x e$

dimana : $D_y = diag(S_{y1}, S_{y2}, \ldots, S_{yp})$

$D_x = diag(S_{x1}, S_{x2}, \ldots, S_{xq})$

13.2.3. Perhitungan Proporsi Keragaman

Besarnya nilai proporsi keragaman menunjukkan baik tidaknya peubah kanonik yang dipilih untuk menerangkan keragaman asal. Semakin besar nilai proporsi keragaman maka semakin baik peubah-peubah kanonik yang dipilih menerangkan keragaman asal. Batasan yang digunakan untuk nilai proporsi bersifat relatif, sebagai acuan lebih besar dari 70%.

13.2.4. Uji Hipotesis

Ada dua hipotesis yang akan diujikan dalam analisis korelasi kanonik yaitu uji hipotesis untuk mengetahui apakah secara keseluruhan korelasi kanonik signifikan (uji korelasi kanonik secara bersama) dan uji hipotesis untuk mengetahui apakah ada sebagian korelasi kanonik signifikan (uji individu). Jika uji hipotesis pertama memperoleh kesimpulan bahwa paling tidak ada ada satu korelasi kanonik tidak bernilai nol maka dilanjutkan dengan uji hipotesis kedua untuk mengetahui apakah ada sebagian korelasi kanonik signifikan.

1. Uji korelasi kanonik secara bersama

Hipotesis :

$H_0 = \rho_1 = \rho_2 = ... = \rho_k = 0$ (semua korelasi kanonik bernilai nol)

H_1 = ada $\rho_i \neq 0$ (paling tidak ada satu korelasi kanonik tidak bernilai nol)

Statistik uji : $B = -[n - 1 - \frac{1}{2}(p + q + 1)]\ln \Lambda$

$$\Lambda = \prod_{i=1}^{k}\left(1 - \rho_i^2\right)$$

dengan : dan n = jumlah pengamatan

Kriteria keputusan : hipotesis nol ditolak pada taraf signifikansi α jika $B > \chi^2_\alpha$ dengan derajat bebas p x q.

2. Uji individu

Hipotesis :

$H_0 = \rho_1 = 0$, $\rho_2 = 0$, ..., $\rho_k = 0$

$H_1 = \rho_i \neq 0$, untuk $i = 1, 2, ..., k$

Statistik Uji :

$$B_r = -[n - 1 - \frac{1}{2}(p + q + 1)]\ln \Lambda_r$$

dengan :

$$\Lambda_r = \prod_{i=r}^{k}\left(1 - \rho_i^2\right)$$

dan n = jumlah pengamatan

Kriteria keputusan : hipotesis nol ditolak pada taraf signifikansi α jika $B_r > \chi^2_\alpha$ dengan derajat bebas $(p-r)$ $(q-r)$.

13.2.5. Interpretasi Fungsi Kanonik

Interpretasi yang dapat dilakukan dalam analisis korelasi kanonik yaitu terhadap koefisien kanonik (bobot kanonik / *weight* kanonik), *loadings* kanonik dan *cross loadings* kanonik.

1. ***Weight* kanonik** merupakan koefisien kanonik yang telah dibakukan, dapat diinterpretasikan sebagai besarnya kontribusi peubah asal terhadap peubah kanonik. Semakin besar nilai koefisien ini maka semakin besar kontribusi peubah yang bersangkutan terhadap peubah kanonik.
2. ***Loadings* kanonik** dapat dihitung dari korelasi antara peubah asal dengan masing-masing peubah kanoniknya. Semakin besar nilai *loading* mencerminkan semakin dekat hubungan fungsi kanonik yang bersangkutan dengan peubah asal. *Loadings* kanonik dibedakan menjadi :
   a. ***Loadings kanonik peubah independen:***

   $$R_{XW} = R_{XX} A_Z$$

   b. ***Loadings kanonik peubah dependen:***

   $$R_{YV} = R_{YY} B_Z$$

   c. ***Cross loadings* kanonik** dapat dihitung dari korelasi antara peubah asal dengan bukan peubah kanoniknya. Semakin besar nilai *loading* mencerminkan semakin dekat hubungan fungsi kanonik yang bersangkutan dengan peubah asal. *Cross loading* kanonik terdiri dari :
      1) ***Cross Loadings kanonik peubah independen:***

      $$\mathbf{R_{XV} = R_{XW}\, \rho_k}$$

      2) ***Cross Loadings kanonik peubah dependen:***

      $$\mathbf{R_{YW} = R_{YV}\, \rho_k}$$

13.2.6. Redundansi

Redundansi adalah suatu indeks yang menghitung proporsi keragaman yang dapat dijelaskan oleh peubah kanonik yang dipilih, baik dari peubah kanonik dependen maupun peubah kanonik independen. Proporsi keragaman dapat dibedakan menjadi:

1. **Proporsi keragaman Y** yang diterangkan oleh peubah kanonik V :

$$R^2_{Z^{(2)}|V_1,V_2,\ldots,V_r} = \frac{\sum_{i=1}^{r} \sum_{k=1}^{q} r^2_{V_i, z_k^{(2)}}}{q}$$

2. **Proporsi keragaman Y** yang diterangkan oleh peubah kanonik W :

$$R^2_{(k)Y|X} = \rho_k^2 . R^2_{(k)Y}$$

3. **Proporsi keragaman X** yang diterangkan oleh peubah kanonik W :

$$R^2_{Z^{(1)}|W_1,W_2,\ldots,W_r} = \frac{\sum_{i=1}^{r} \sum_{k=1}^{p} r^2_{W_i, z_k^{(1)}}}{p}$$

4. **Proporsi keragaman X** yang diterangkan oleh peubah kanonik V :

$$R^2_{(k)X|Y} = \rho_k^2 . R^2_{(k)X}$$

Koefisien R-square digunakan untuk menentukan fungsi kanonik yang dianggap cukup untuk menerangkan struktur hubungan Y dan X. Nilai ini diperoleh dengan mengkuadratkan korelasi kanonik yang dinotasikan sebagai berikut :

$$R_k^2 = \rho_k^2$$

13.2.7. Validasi Fungsi Kanonik

Validasi perlu dilakukan sebelum mengambil kesimpulan untuk kasus yang dihadapi. Tahap ini dilakukan untuk meyakinkan bahwa hasil yang telah diperoleh sebelumnya bersifat umum. Atau dengan kata lain, tahap ini dapat mengidentifikasikan tingkat kestabilan dari fungsi kanonik yang diperoleh. Jika *loading* kanonik stabil maka dapat dikatakan bahwa fungsi kanonik yang diperoleh sudah valid unuk menarik kesimpulan.

Berikut ini adalah beberapa metode validasi yang dapat digunakan :

1. Membagi sampel menjadi dua bagian, bagian pertama digunakan untuk menduga fungsi kanonik dan bagian kedua digunakan sebagai validasi.
2. Analisis sensitivitas untuk peubah-peubah independen, yaitu dengan membandingkan loading kanonik apabila salah-satu dari peubah independen disisihkan dari analisis.

13.3. Ilustrasi

Survey HATCO dilakukan terhadap 100 responden yang merupakan manajer pembelian dari perusahaan-perusahaan yang membeli barang di HATCO. Tujuan survey ini adalah ingin mengetahui persepsi terhadap perusahaan HATCO sebagai perusahaan pemasok (*supplier*) berdasarkan hasil pembelian. Peubah-peubah yang menggambarkan **persepsi terhadap perusahaan** telah diidentifikasi pada studi sebelumnya terdiri dari 7 peubah yang paling berpengaruh yaitu :

X1 : Kecepatan Pengantaran (waktu yang dibutuhkan untuk mengantarkan produk setelah perintah telah dikonfirmasi)

X2 : Tingkat Harga

X3 : Fleksibilitas Harga (persepsi mengenai kesediaan perwakilan HATCO untuk menegosiasikan harga pada semua jenis pembelian)

X4 : Citra Pabrik Pembuat (gambaran secara menyeluruh mengenai perusahaan HATCO sebagai *supplier*)

X5 : Layanan Keseluruhan (keseluruhan tingkat layanan yang diperlukan untuk mempertahankan hubungan yang memuaskan antara pemasok dan pembeli)

X6 : Citra Tenaga Penjual

X7 : Kualitas Produk (persepsi tingkat kualitas produk tertentu misalnya: kinerja atau hasil)

Peubah X1-X7 dalam selang kontinu 0 s.d. 10 dimana 0 = buruk dan 10 = sangat baik

Hasil pembelian terdiri dari 2 peubah yang diperoleh dari hasil evaluasi kepuasan masing-masing responden terhadap HATCO dan persentase pembelian produk yang dibeli responden yaitu :

X9 : Tingkat Penggunaan (berapa banyak produk total perusahaan yang dibeli dari HATCO, diukur berdasarkan persentase dengan skala 100, mulai dari 0 hingga 100 persen

X10 : Tingkat Kepuasan (seberapa puas pembeli dengan pembelian terakhir dari HATCO, diukur dengan rating skala seperti pada X1 untuk X7)

Data selengkapnya adalah

ID Responden	Peubah Independen								Peubah Dependen	
	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10
1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**
1,0	4,1	0,6	6,9	4,7	2,4	2,3	5,2	0	32,0	4,2
2,0	1,8	3,0	6,3	6,6	2,5	4,0	8,4	1	43,0	4,3
3,0	3,4	5,2	5,7	6,0	4,3	2,7	8,2	1	48,0	5,2
4,0	2,7	1,0	7,1	5,9	1,8	2,3	7,8	1	32,0	3,9
5,0	6,0	0,9	9,6	7,8	3,4	4,6	4,5	0	58,0	6,8
6,0	1,9	3,3	7,9	4,8	2,6	1,9	9,7	1	45,0	4,4
7,0	4,6	2,4	9,5	6,6	3,5	4,5	7,6	0	46,0	5,8
8,0	1,3	4,2	6,2	5,1	2,8	2,2	6,9	1	44,0	4,3
9,0	5,5	1,6	9,4	4,7	3,5	3,0	7,6	0	63,0	5,4
10,0	4,0	3,5	6,5	6,0	3,7	3,2	8,7	1	54,0	5,4
11,0	2,4	1,6	8,8	4,8	2,0	2,8	5,8	0	32,0	4,3
12,0	3,9	2,2	9,1	4,6	3,0	2,5	8,3	0	47,0	5,0
13,0	2,8	1,4	8,1	3,8	2,1	1,4	6,6	1	39,0	4,4
14,0	3,7	1,5	8,6	5,7	2,7	3,7	6,7	0	38,0	5,0
15,0	4,7	1,3	9,9	6,7	3,0	2,6	6,8	0	54,0	5,9
16,0	3,4	2,0	9,7	4,7	2,7	1,7	4,8	0	49,0	4,7
17,0	3,2	4,1	5,7	5,1	3,6	2,9	6,2	0	38,0	4,4
18,0	4,9	1,8	7,7	4,3	3,4	1,5	5,9	0	40,0	5,6
19,0	5,3	1,4	9,7	6,1	3,3	3,9	6,8	0	54,0	5,9
20,0	4,7	1,3	9,9	6,7	3,0	2,6	6,8	0	55,0	6,0
21,0	3,3	0,9	8,6	4,0	2,1	1,8	6,3	0	41,0	4,5
22,0	3,4	0,4	8,3	2,5	1,2	1,7	5,2	0	35,0	3,3
23,0	3,0	4,0	9,1	7,1	3,5	3,4	8,4	0	55,0	5,2
24,0	2,4	1,5	6,7	4,8	1,9	2,5	7,2	1	36,0	3,7
25,0	5,1	1,4	8,7	4,8	3,3	2,6	3,8	0	49,0	4,9
26,0	4,6	2,1	7,9	5,8	3,4	2,8	4,7	0	49,0	5,9
27,0	2,4	1,5	6,6	4,8	1,9	2,5	7,2	1	36,0	3,7
28,0	5,2	1,3	9,7	6,1	3,2	3,9	6,7	0	54,0	5,8
29,0	3,5	2,8	9,9	3,5	3,1	1,7	5,4	0	49,0	5,4

ID Responden	Peubah Independen								Peubah Dependen	
	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10
30,0	4,1	3,7	5,9	5,5	3,9	3,0	8,4	1	46,0	5,1
31,0	3,0	3,2	6,0	5,3	3,1	3,0	8,0	1	43,0	3,3
32,0	2,8	3,8	8,9	6,9	3,3	3,2	8,2	0	53,0	5,0
33,0	5,2	2,0	9,3	5,9	3,7	2,4	4,6	0	60,0	6,1
34,0	3,4	3,7	6,4	5,7	3,5	3,4	8,4	1	47,0	3,8
35,0	2,4	1,0	7,7	3,4	1,7	1,1	6,2	1	35,0	4,1
36,0	1,8	3,3	7,5	4,5	2,5	2,4	7,6	1	39,0	3,6
37,0	3,6	4,0	5,8	5,8	3,7	2,5	9,3	1	44,0	4,8
38,0	4,0	0,9	9,1	5,4	2,4	2,6	7,3	0	46,0	5,1
39,0	0,0	2,1	6,9	5,4	1,1	2,6	8,9	1	29,0	3,9
40,0	2,4	2,0	6,4	4,5	2,1	2,2	8,8	1	28,0	3,3
41,0	1,9	3,4	7,6	4,6	2,6	2,5	7,7	1	40,0	3,7
42,0	5,9	0,9	9,6	7,8	3,4	4,6	4,5	0	58,0	6,7
43,0	4,9	2,3	9,3	4,5	3,6	1,3	6,2	0	53,0	5,9
44,0	5,0	1,3	8,6	4,7	3,1	2,5	3,7	0	48,0	4,8
45,0	2,0	2,6	6,5	3,7	2,4	1,7	8,5	1	38,0	3,2
46,0	5,0	2,5	9,4	4,6	3,7	1,4	6,3	0	54,0	6,0
47,0	3,1	1,9	10,0	4,5	2,6	3,2	3,8	0	55,0	4,9
48,0	3,4	3,9	5,6	5,6	3,6	2,3	9,1	1	43,0	4,7
49,0	5,8	0,2	8,8	4,5	3,0	2,4	6,7	0	57,0	4,9
50,0	5,4	2,1	8,0	3,0	3,8	1,4	5,2	0	53,0	3,8
51,0	3,7	0,7	8,2	6,0	2,1	2,5	5,2	0	41,0	5,0
52,0	2,6	4,8	8,2	5,0	3,6	2,5	9,0	1	53,0	5,2
53,0	4,5	4,1	6,3	5,9	4,3	3,4	8,8	1	50,0	5,5
54,0	2,8	2,4	6,7	4,9	2,5	2,6	9,2	1	32,0	3,7
55,0	3,8	0,8	8,7	2,9	1,6	2,1	5,6	0	39,0	3,7
56,0	2,9	2,6	7,7	7,0	2,8	3,6	7,7	0	47,0	4,2
57,0	4,9	4,4	7,4	6,9	4,6	4,0	9,6	1	62,0	6,2
58,0	5,4	2,5	9,6	5,5	4,0	3,0	7,7	0	65,0	6,0
59,0	4,3	1,8	7,6	5,4	3,1	2,5	4,4	0	46,0	5,6
60,0	2,3	4,5	8,0	4,7	3,3	2,2	8,7	1	50,0	5,0
61,0	3,1	1,9	9,9	4,5	2,6	3,1	3,8	0	54,0	4,8
62,0	5,1	1,9	9,2	5,8	3,6	2,3	4,5	0	60,0	6,1
63,0	4,1	1,1	9,3	5,5	2,5	2,7	7,4	0	47,0	5,3
64,0	3,0	3,8	5,5	4,9	3,4	2,6	6,0	0	36,0	4,2
65,0	1,1	2,0	7,2	4,7	1,6	3,2	10,0	1	40,0	3,4

ID Responden	Peubah Independen								Peubah Dependen	
	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10
66,0	3,7	1,4	9,0	4,5	2,6	2,3	6,8	0	45,0	4,9
67,0	4,2	2,5	9,2	6,2	3,3	3,9	7,3	0	59,0	6,0
68,0	1,6	4,5	6,4	5,3	3,0	2,5	7,1	1	46,0	4,5
69,0	5,3	1,7	8,5	3,7	3,5	1,9	4,8	0	58,0	4,3
70,0	2,3	3,7	8,3	5,2	3,0	2,3	9,1	1	49,0	4,8
71,0	3,6	5,4	5,9	6,2	4,5	2,9	8,4	1	50,0	5,4
72,0	5,6	2,2	8,2	3,1	4,0	1,6	5,3	0	55,0	3,9
73,0	3,6	2,2	9,9	4,8	2,9	1,9	4,9	0	51,0	4,9
74,0	5,2	1,3	9,1	4,5	3,3	2,7	7,3	0	60,0	5,1
75,0	3,0	2,0	6,6	6,6	2,4	2,7	8,2	1	41,0	4,1
76,0	4,2	2,4	9,4	4,9	3,2	2,7	8,5	0	49,0	5,2
77,0	3,8	0,8	8,3	6,1	2,2	2,6	5,3	0	42,0	5,1
78,0	3,3	2,6	9,7	3,3	2,9	1,5	5,2	0	47,0	5,1
79,0	1,0	1,9	7,1	4,5	1,5	3,1	9,9	1	39,0	3,3
80,0	4,5	1,6	8,7	4,6	3,1	2,1	6,8	0	56,0	5,1
81,0	5,5	1,8	8,7	3,8	3,6	2,1	4,9	0	59,0	4,5
82,0	3,4	4,6	5,5	8,2	4,0	4,4	6,3	0	47,0	5,6
83,0	1,6	2,8	6,1	6,4	2,3	3,8	8,2	1	41,0	4,1
84,0	2,3	3,7	7,6	5,0	3,0	2,5	7,4	0	37,0	4,4
85,0	2,6	3,0	8,5	6,0	2,8	2,8	6,8	1	53,0	5,6
86,0	2,5	3,1	7,0	4,2	2,8	2,2	9,0	1	43,0	3,7
87,0	2,4	2,9	8,4	5,9	2,7	2,7	6,7	1	51,0	5,5
88,0	2,1	3,5	7,4	4,8	2,8	2,3	7,2	0	36,0	4,3
89,0	2,9	1,2	7,3	6,1	2,0	2,5	8,0	1	34,0	4,0
90,0	4,3	2,5	9,3	6,3	3,4	4,0	7,4	0	60,0	6,1
91,0	3,0	2,8	7,8	7,1	3,0	3,8	7,9	0	49,0	4,4
92,0	4,8	1,7	7,6	4,2	3,3	1,4	5,8	0	39,0	5,5
93,0	3,1	4,2	5,1	7,8	3,6	4,0	5,9	0	43,0	5,2
94,0	1,9	2,7	5,0	4,9	2,2	2,5	8,2	1	36,0	3,6
95,0	4,0	0,5	6,7	4,5	2,2	2,1	5,0	0	31,0	4,0
96,0	0,6	1,6	6,4	5,0	0,7	2,1	8,4	1	25,0	3,4
97,0	6,1	0,5	9,2	4,8	3,3	2,8	7,1	0	60,0	5,2
98,0	2,0	2,8	5,2	5,0	2,4	2,7	8,4	1	38,0	3,7
99,0	3,1	2,2	6,7	6,8	2,6	2,9	8,4	1	42,0	4,3
100,0	2,5	1,8	9,0	5,0	2,2	3,0	6,0	0	33,0	4,4

Hasil Analisis

Berikut adalah beberapa hasil yang diperoleh dalam analisis korelasi kanonik (diperoleh menggunakan program SAS).

1. Nilai Korelasi Pearson Antar Peubah Dependen, Antar Peubah Independen Dan Korelasi Silang Antara Peubah Dependen Dengan Independen

Hasil Analisis Korelasi Kanonik

The CANCORR Procedure

Correlations Among the Original Variables

Correlations Among the VAR Variables

	X9	X10
X9	1.0000	0.7107
X10	0.7107	1.0000

Correlations Among the WITH Variables

	X1	X2	X3	X4	X5	X6	X7
X1	1.0000	-0.3492	0.5093	0.0504	0.6119	0.0771	-0.4826
X2	-0.3492	1.0000	-0.4872	0.2722	0.5130	0.1862	0.4697
X3	0.5093	-0.4872	1.0000	-0.1161	0.0666	-0.0343	-0.4481
X4	0.0504	0.2722	-0.1161	1.0000	0.2987	0.7882	0.2000
X5	0.6119	0.5130	0.0666	0.2987	1.0000	0.2408	-0.0552
X6	0.0771	0.1862	-0.0343	0.7882	0.2408	1.0000	0.1773
X7	-0.4826	0.4697	-0.4481	0.2000	-0.0552	0.1773	1.0000

Correlations Between the VAR Variables and the WITH Variables

	X1	X2	X3	X4	X5	X6	X7
X9	0.6765	0.0819	0.5590	0.2242	0.7007	0.2561	-0.1925
X10	0.6506	0.0284	0.5248	0.4759	0.6312	0.3409	-0.2833

Dari nilai korelasi antar peubah dependen terlihat bahwa ada hubungan linear yang erat antara Tingkat Penggunaan (X9) dan Tingkat Kepuasan (X10). Sedangkan korelasi antar peubah independen terlihat bahwa ada hubungan linear yang erat antara Citra Pabrik Pembuat(X4) dengan Citra Tenaga Penjual(X6). Hubungan linear yang paling rendah terjadi antara Fleksibilitas Harga (X3) dengan Citra Tenaga Penjual(X6).

Korelasi antar peubah dependen dengan peubah independen terlihat bahwa peubah Tingkat Penggunaan (X9) dan Tingkat Kepuasan (X10) memiliki hubungan linear yang cukup erat dengan Kecepatan Pengantaran (X1) dan Layanan Keseluruhan (X5). Sedangkan hubungan antar peubah yang lainnya terlihat tingkat hubungan linearnya relatif rendah.

2. Korelasi Kanonik

   Korelasi kanonik dari gugus peubah dependen dengan peubah independen menghasilkan 2 fungsi kanonik. Korelasi kanonik pertama sebesar 0.937 dan kedua sebesar 0.51. Dari hasil tersebut menunjukkan bahwa korelasi kanonik yang cukup besar terjadi pada fungsi kanonik pertama.

Canonical Correlation Analysis

Squared Canonical Correlation		Canonical Correlation	Adjusted Canonical Correlation	Approximate Standard Error
	1	0.936913	0.932812	0.012281

3. Inferensia

   Untuk memilih jumlah fungsi kanonik yang cukup digunakan untuk menjelaskan hubungan antara gugus peubah dependen dengan independen dapat dilihat dari kontribusi keragaman kumulatif yag dijelaskan oleh fungsi kanonik atau berdasarkan uji statistik.

   Berdasarkan kontribusi keragaman yang dijelaskan oleh fungsi kanonik terlihat bahwa fungsi kanonik pertama menjelaskan keragaman total sebesar 95.33% sedangkan yang kedua sebesar 4.67%. Hasil ini menunjukkan bahwa untuk menerangkan keragaman total cukup mengambil fungsi kanonik pertama saja.

Berdasarkan uji rasio kemungkinan terlihat bahwa korelasi kanonik pertama dan kedua berbeda nyata pada a=0.05. Hal ini berarti korelasi kanonik yang dapat digunakan untuk menjelaskan hubungan antar gugus peubah dependen dan independen adalah dua korelasi kanonik.

Pengujian korelasi kanonik secara simultan menunjukkan bahwa paling sedikit ada satu korelasi kanonik yang nyata. Dengan demikian dari kedua metode pengujian diatas dapat disimpulkan bahwa korelasi kanonik yang dipilih adalah dua korelasi kanonik.

Test of H0: The canonical correlations in the current row and all that follow are zero

Eigenvalues of Inv(E)*H = CanRsq/(1-CanRsq)

	Eigenvalue	Difference	Proportion	Cumulative	Likelihood Ratio	Approximate F Value	Num DF	Den DF	Pr > F
1	7.1837	6.8321	0.9533	0.9533	0.09040876	30.24	14	182	<.0001
2	0.3516		0.0467	1.0000	0.73987709	5.39	6	92	<.0001

Multivariate Statistics and F Approximations

S=2 M=2 N=44.5

Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.09040876	30.24	14	182	<.0001
Pillai's Trace	1.13792861	17.35	14	184	<.0001
Hotelling-Lawley Trace	7.53526400	48.59	14	142.28	<.0001
Roy's Greatest Root	7.18368817	94.41	7	92	<.0001

NOTE: F Statistic for Roy's Greatest Root is an upper bound.

NOTE: F Statistic for Wilks' Lambda is exact.

4. Bentuk Kedua Fungsi Kanonik

   Berdasarkan koefisien kanonik yang telah dibakukan untuk fungsi kanonik dependen terlihat bahwa urutan kontribusi relatif peubah-peubah dependen terhadap variate pertama adalah X10 dan X9. Urutan kontribusi relatif peubah-peubah dependen terhadap variate kedua adalah X9 dan X10.

   Berdasarkan koefisien kanonik yang telah dibakukan untuk fungsi kanonik independen terlihat bahwa urutan kontribusi relatif peubah-peubah independen terhadap variate pertama adalah

   X3,X5,X4,X1,X2,X6 dan X7. Urutan kontribusi relatif peubah-peubah dependen terhadap variate kedua adalah X5,X4,X1,X2,X6,X7dan X3.

```
              Canonical Correlation Analysis
   Raw Canonical Coefficients for the VAR Variables
                                    V1                V2
                X9       0.0556978517      0.1480049728
                X10      0.6780451833      -1.516799589
            Raw Canonical Coefficients for the WITH Variables
                                  W1                W2
                X1      0.1702660392      -0.730470504
                X2      0.0860216803      -0.725990627
                X3      0.4100729309      0.1151226504
                X4      0.3075988599      -1.286602814
                X5      0.5927979812      2.0371773018
                X6      -0.065847588      0.9550978687
                X7      0.0004309902       0.301249326

              Canonical Correlation Analysis
  Standardized Canonical Coefficients for the VAR Variables
                                    V1            V2
                    X9          0.5007        1.3304
                    X10         0.5801       -1.2977

 Standardized Canonical Coefficients for the WITH Variables
                                  W1            W2
                    X1         0.2249       -0.9648
                    X2         0.1029       -0.8680
                    X3         0.5686        0.1596
                    X4         0.3480       -1.4557
                    X5         0.4453        1.5304
                    X6        -0.0508        0.7362
                    X7         0.0007        0.4776
```

5. Korelasi Antara Setiap Peubah Dengan Fungsi Kanoniknya (*Loading* Kanonik)

Untuk fungsi kanonik peubah-peubah dependen terlihat bahwa peubah yang berhubungan paling erat dengan fungsi kanonik pertama adalah X9 dan X10 sedangkan terhadap fungsi kanonik kedua yaitu X9. Untuk fungsi kanonik peubah independen terlihat bahwa peubah-peubah yang berhubungan paling erat dengan fungsi kanonik pertama adalah X1,X3 dan X5 sedangkan terhadap fungsi kanonik kedua yaitu X4.

```
                              Canonical Structure

                 Correlations Between the VAR Variables and Their Canonical Variables
                                           V1             V2
                                  X9     0.9129         0.4081
                                  X10    0.9359        -0.3522

                Correlations Between the WITH Variables and Their Canonical Variables
                                          W1             W2
                                  X1    0.7643         0.1091
                                  X2    0.0614         0.1414
                                  X3    0.6237         0.1229
                                  X4    0.4145        -0.6262
                                  X5    0.7653         0.2216
                                  X6    0.3479        -0.1995
                                  X7   -0.2783         0.2189

Correlations Between the VAR Variables and the Canonical Variables of the WITH Variables
                                           W1             W2
                                  X9     0.8553         0.2081
                                  X10    0.8769        -0.1796

Correlations Between the WITH Variables and the Canonical Variables of the VAR Variables
                                          V1             V2
                                  X1    0.7161         0.0556
                                  X2    0.0575         0.0721
                                  X3    0.5843         0.0627
                                  X4    0.3883        -0.3194
                                  X5    0.7170         0.1130
                                  X6    0.3260        -0.1017
                                  X7   -0.2607         0.1116
```

Dari korelasi silang antar peubah-peubah dependen terhadap fungsi kanonik peubah independen terlihat bahwa yang berhubungan paling erat dengan fungsi kanonik pertama adalah X9 dan X10 serta terhadap fungsi kanonik kedua adalah X9.

Dari korelasi silang antar peubah-peubah independen terhadap fungsi kanonik peubah dependen terlihat bahwa yang

berhubungan paling erat dengan fungsi kanonik pertama adalah X1 dan X5 serta terhadap fungsi kanonik kedua adalah X4.

6. Redudansi

Dari hasil analisis redudansi terlihat bahwa cukup korelasi kanonik pertama saja yang digunakan karena R^2 kanonik untuk fungsi kanonik kedua kecil yaitu 26.01%.

Canonical Redundancy Analysis

Raw Variance of the VAR Variables Explained by

Canonical Variable Number	Their Own Canonical Variables			The Opposite Canonical Variables	
	Proportion	Cumulative Proportion	Canonical R-Square	Proportion	Cumulative Proportion
1	0.8338	0.8338	0.8778	0.7319	0.7319
2	0.1662	1.0000	0.2601	0.0432	0.7752

Raw Variance of the WITH Variables Explained by

Canonical Variable Number	Their Own Canonical Variables			The Opposite Canonical Variables	
	Proportion	Cumulative Proportion	Canonical R-Square	Proportion	Cumulative Proportion
1	0.2577	0.2577	0.8778	0.2262	0.2262
2	0.0748	0.3325	0.2601	0.0195	0.2457

Canonical Redundancy Analysis

Standardized Variance of the VAR Variables Explained by

Canonical Variable Number	Their Own Canonical Variables			The Opposite Canonical Variables	
	Proportion	Cumulative Proportion	Canonical R-Square	Proportion	Cumulative Proportion
1	0.8547	0.8547	0.8778	0.7503	0.7503
2	0.1453	1.0000	0.2601	0.0378	0.7881

Standardized Variance of the WITH Variables Explained by

Canonical Variable Number	Their Own Canonical Variables			The Opposite Canonical Variables	
	Proportion	Cumulative Proportion	Canonical R-Square	Proportion	Cumulative Proportion
1	0.2761	0.2761	0.8778	0.2424	0.2424
2	0.0823	0.3584	0.2601	0.0214	0.2638

7. Validasi Hasil

Validasi hasil untuk kasus ini menggunakan metode analisis sensitivitas untuk peubah-peubah independen, yaitu dengan membandingkan loading kanonik apabila salah-satu dari peubah independen disisihkan dari analisis. Hasil analisis sebelumnya akan valid jika penyisihan salah satu peubah independen tersebut tetap menghasilkan korelasi kanonik yang stabil.

Banyaknya peubah dalam kasus ini menyebabkan pilihan peubah yang dapat disisihkan sangat banyak. Hair,et.al (1995) melakukan penyisihan peubah pada X1, X2 dan X7 dan menyatakan tidak terjadi perubahan kekuatan dan koefisien korelasi kanonik. Selain itu menyimpulkan juga bahwa tiga peubah dalam kelompok persepsi terhadap perusahaan yang terpenting adalah X1,X3 dan X5.

Dalam tulisan ini,uji sensitivitas akan dilakukan dengan menyisihkan 3 peubah yaitu X1, X2 dan X5. Hasil yang dimuat dalam Tabel 1 menunjukkan adanya stabilitas model korelasi kanonik serta dengan mengikuti analisis Hair,et.al (1995), analisis ini juga

menunjukkan bahwa tiga peubah dalam kelompok persepsi terhadap perusahaan yang paling penting adalah X1,X3 dan X5.

Tabel 1 Hasil Analisis Sensitivitas Korelasi Kanonik

	Complete Variate	Hasil Setelah Penyisihan peubah X1	X2	X5
1	**2**	**3**	**4**	**5**
Korelasi Kanonik (*R*)	0.937	0.936	0.937	0.934838
Root Kanonik (R^2)	0.878	0.877	0.878	0.873922
Peubah Independen				
X1 Kecepatan Pengantaran	0.764	dihilangkan	0.765	0.7653
X2 Tingkat Harga	0.061	0.062	dihilangkan	0.0606
X3 Fleksibilitas Harga	0.624	0.624	0.624	0.6242
X4 Citra Pabrik Pembuat	0.414	0.413	0.414	0.4194
X5 Layanan Keseluruhan	0.765	0.766	0.766	dihilangkan
X6 Citra Tenaga Penjual	0.348	0.348	0.348	0.3499
X7 Kualitas Produk	-0.278	-0.278	-0.278	-0.2803
shared variance	0.2761	0.225	0.3216	0.2260
Redudansi	0.2424	0.1972	0.2822	0.1975
Peubah dependen				
Loading kanonik				
X9 Tingkat Penggunaan	0.913	0.915	0.914	0.9080
X10 Tingkat Kepuasan	0.936	0.934	0.935	0.9400
shared variance	0.8547	0.8549	0.8548	0.8541
Redudansi	0.7503	0.7479	0.7501	0.7464

Berdasarkan hasil analisis di atas, persepsi terhadap perusahaan HATCO khususnya kecepatan pengantaran, fleksibilitas harga dan layanan berhubungan erat dengan hasil pembelian dilihat dari tingkat kepuasaan dan tingkat penggunaan.

13.4. Aplikasi SAS

Berikut adalah program SAS dari data seperti pada ilustrasi di atas.

```
DATA HATCO;
INPUT ID X1 X2 X3 X4 X5 X6 X7 X8 X9 X10 X11 X12 X13 X14;
DATALINES;
1    4.1  0.6  6.9  4.7  2.4  2.3  5.2  0  32  4.2  1  0  1  1
2    1.8  3    6.3  6.6  2.5  4    8.4  1  43  4.3  0  1  0  1
3    3.4  5.2  5.7  6    4.3  2.7  8.2  1  48  5.2  0  1  1  2
4    2.7  1    7.1  5.9  1.8  2.3  7.8  1  32  3.9  0  1  1  1
5    6    0.9  9.6  7.8  3.4  4.6  4.5  0  58  6.8  1  0  1  3
⋮
97   6.1  0.5  9.2  4.8  3.3  2.8  7.1  0  60  5.2  1  0  1  3
98   2    2.8  5.2  5    2.4  2.7  8.4  1  38  3.7  0  1  0  1
99   3.1  2.2  6.7  6.8  2.6  2.9  8.4  1  42  4.3  0  1  0  1
100  2.5  1.8  9    5    2.2  3    6    0  33  4.4  1  0  0  1

;

Title 'Hasil Analisis Korelasi Kanonik';

Proc cancorr REDUNDANCY corr data=HATCO;

var X9 X10;

with X1 X2 X3 X4 X5 X6 X7;

run;
```

14

14. Analisis Regresi Peubah Ganda (*Multivariate Regression Analysis*)

14.1. Pendahuluan

Pada saat sekarang, perkembangan penelitian pada bidang ekonometrika, pendidikan, psikometrika, dan beberapa bidang ilmu pengetahuan sosial lainnya, tertarik pada pemodelan hubungan yang relatif rumit. Pemodelan hubungan tersebut dibentuk melalui satu atau lebih peubah tak bebas yang dijelaskan oleh satu atau lebih peubah bebas, dan pada saat yang bersamaan satu atau lebih peubah tak bebas tersebut berperan sebagai peubah bebas bagi peubah tak bebas lainnya.

Peubah-peubah pada model hubungan yang relatif rumit tersebut dapat berupa peubah terukur maupun peubah tak terukur (peubah laten). Peubah terukur pada model hubungan tersebut dapat langsung diketahui nilai/besarnya melalui suatu pengamatan, sedangkan untuk peubah laten nilai/besarnya dibangun melalui beberapa peubah *manifest* (penjelas) yang diasumsikan merupakan indikator (pengukur) bagi peubah laten.

Model hubungan yang relatif rumit dengan sebagian atau seluruh peubah-peubahnya berupa peubah laten, menyulitkan peneliti dalam menganalisis keterkaitan hubungan antar peubah-peubah laten

tersebut. Kesulitan ini disebabkan karena pendugaan terhadap model hubungan antar peubah laten dan pengujian terhadap model dugaannya tidak dapat dilakukan menggunakan analisis regresi secara simultan (pada waktu yang bersamaan). Kesulitan inilah yang menuntun peran statistikawan untuk mengembangkan suatu teknik analisis yang dapat membantu dalam menyelesaikan permasalahan tersebut. Oleh karena itu, banyak analisis statistika yang berkembang pesat untuk mengatasi hal tersebut yaitu salah satunya dengan menggunakan analisis regresi multivariate yang mana dalam regresi ini terdapat beberapa variabel y dan beberapa variabel x dimana variabel y saling berhubungan sehingga harus terdapat besarnya korelasi antara variabel y tersebut untuk menentukan persamaan structural yang dapat menginformasikan secara lebih spesifik.

14.2. Analisis Regresi

Berdasarkan jumlah variable bebas dan tak bebas yang dilibatkan dalam analisis regresi, analisis regresi dapat dibedakan menjadi analisis regresi linear sederhana, analisis regresi linear berganda, dan analisis regresi linear multivariate. Secara umum gambaran mengenai perbedaan regresi linear, regresi berganda dan regresi multivariate dapat di lihat sbb :

- Regresi linier sederhana : hanya terdiri dari satu y dan satu x. sebagai contoh : dalam suatu universitas meramalkan IPK mahasiswa berdasarkan pada nilai-nilai yang ada di SMA.
- Regresi linier berganda : terdiri dari satu y dan beberapa x. Kita dapat meningkatkan prediksi IPK mahasiswa menggunakan lebih dari satu variabel x. Contohnya nilai SMA, standar nilai, peringkat SMA.

- Regresi linier multivariate : terdiri beberapa y dan beberapa x. pada ilutrasi di atas, kita dapat menduga y. tetapi dalam hal ini, variabel y dapat diregresikan kembali dengan variabel x.

14.3. Analisis Regresi Multivariate

Analisis multivariat adalah analisis statistika yang dikenakan pada data yang terdiri dari banyak variable dan antar variable saling berkorelasi. Data multivariat tidak hanya terdiri dari satu variable saja melainkan dapat terdiri atas lebih dari satu variabel. Misal data dari n pengamatan pada p variabel sehingga dapat disusun matriks dengan n baris dan p kolom, dinotasikan X sbb (Morrison, 2005) :

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \cdots & \vdots \\ x_{n1} & x_{n2} & & x_{np} \end{bmatrix}$$

Model regresi multivariat adalah model regresi dengan lebih dari satu variabel respon yang saling berkorelasi dan satu atau lebih variabel predictor (Johnson dan Wichern, 2007; Rencher 2002). Terdapat 2 tipe dasar variabel x pada regresi multivariate yaitu x yang tetap dan x yang random/acak. Pada ilustrasi di atas, variabel x adalah bersifat random karena semuanya tidak dapat di kontrol oleh peneliti. Apabila seseorang memilih x sebagai random, maka semua variabel x dan variabel y dapat diukur dan diamati.

Dalam beberapa kasus percobaan, variabel x adalah bersifat tetap, di bawah pengawasan peneliti. Sebagai contoh, peneliti ingin menghubungkan lahan per hektar dan nilai nutrisi untuk beberapa tingkat pupuk buatan yang diberikan. Peneliti dapat memilih jumlah kimia yang diberikan dan kemudian mengamati perubahan respon

pada lahan dan nutrisi yang diberikan tersebut. Dalam makalah ini, hanya membahas tentang variabel x yang bersifat tetap.

Misalkan terdapat variabel respon berjumlah q yaitu $Y_1, Y_2, \dots, Y_q$ dan p variabel predictor yaitu $X_1, X_2, \dots, X_p$, maka model linear multivariate respon ke-q adalah :

$$Y_1 = \beta_{01} + \beta_{02}X_1 + \cdots + \beta_{p1}X_p + \varepsilon_1$$

$$Y_2 = \beta_{02} + \beta_{12}X_1 + \dots + \beta_{p2}X_p + \varepsilon_2$$

$$\vdots$$

$$Y_q = \beta_{0q} + \beta_{1q}X_1 + \cdots + \beta_{pq}X_p + \varepsilon_q$$

Model regresi multivariate yang terdiri dari q model linear secara simultan dapat ditunjukkan bentuk matriks pada persamaan sbb :

$Y_{(nxq)} = X_{nx(p+1)}\beta_{(p+1)xq} + \varepsilon_{(nxq)}$ $\quad dengan\ asumsi$:

$E(\varepsilon_{(ij)}) = 0\ untuk\ i = 1,2, \dots, n.$

$var(\varepsilon_i) = 0\ untuk\ i = 1,2, \dots, n.$ dan

$Cov(\varepsilon_{(i)}\varepsilon_{(i)}) = \sigma_{ii}I\ , untuk\ i \neq j$

Asumsi 1 menyatakan bahwa model adalah linier dan tidak ada penambahan kondisi yang dibutuhkan untuk menduga y, semua variasi yang tertinggal adalah y yang murni bersifat random dan tidak dapat diduga. Kemudian jika $E(\varepsilon_{(ij)}) = 0$ dan x bersifat tetap, maka $E(y_i) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_q x_{iq}$, dan rata-rata y menggambarkan kondisi dari x tanpa ada penambahan. Asumsi 2, variansi pada setiap ε_i adalah sama yang mana berimplikasi bahwa var(y_i)=σ^2, ketika x bersifat tetap. Asumsi 3 ditimbulkan oleh kondisi yang errornya tidak berkorelasi, dari

pernyataan yang disampaikan menyatakan bahwa y juga tidak berkorelasi yaitu cov (y_i, y_j)=0.

Sehingga 3 asumsi diatas dapat diubah kedalam kondisi y sbb :

$E(y_i) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_q x_{iq}$, i=1,2,….,n.

var(y_i)=σ^2, i=1, 2, …, n.

cov (y_i, y_j)=0, untuk semua i≠j.

Apabila menggunakan notasi matriks, model untuk n observasi dapat dituliskan sbb:

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \left(\begin{pmatrix} 1 & x_{11} & x_{12} \cdots & x_{1q} \\ 1 & x_{21} & x_{22} \cdots & x_{2q} \\ 1 & & \cdots & \\ \vdots & & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} \cdots & x_{nq} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_q \end{pmatrix} + \begin{pmatrix} \varepsilon_0 \\ \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix} \right)$$

Atau dapat dituliskan y=Xβ+ε.

Berdasarkan pada notasi tersebut, 3 asumsi di atas menjadi :

1. E(ε) =0
2. Cov (ε) =σ^2I

Yang mana dapat dituliskan dalam kondisi y sbb :

1. E(y) = Xβ
2. Cov(y) = σ^2I

➢ **Jumlah Kuadrat Sisaan**

Untuk menduga β_0,β_1, … ,β_q yaitu dengan meminimumkan jumlah kuadrat sisaan n pengamatan y dari nilai modelnya, yaitu dari nilai $\hat{y_i}$ di duga menggunakan model. Sehingga dapat dilihat sbb :

$SSE = \sum_{i=1}^{n} \widehat{\varepsilon_i^2} = \sum_{i=1}^{n} \quad (y_i - \widehat{y_i})^2$

$= \sum_{i=1}^{n} \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2} - \cdots - \hat{\beta}_q x_{iq}\right)^2$

Sehingga nilai dugaan β dapat diberikan sbb :

$$\hat{\beta} = (X'X)^{-1}X'y$$

- **Penduga σ^2**

Ditunjukkan bahwa,

$$E(SSE) = \sigma^2[n - (q + 1)] = \sigma^2(n - q - 1)$$

Sehingga penduga tak bias dari σ^2 adalah sebagai berikut :

$$s^2 = \frac{SSE}{n - q - 1} = \frac{1}{n - q - 1}(y - X\hat{\beta})'(y - X\hat{\beta})$$

Dimana SSE dalam bentuk

$$SSE = y'y - \widehat{\beta}'X'y$$

- **Rata-rata Model Terkoreksi**

Terkadang variabel x menggunakan rataan data tersebut. Seperti $\overline{x_1} = \sum_{i=1}^{n} \frac{x_{i1}}{n}, \overline{x_2} = \sum_{i=1}^{n} x_{i2}/n$, sehingga model menjadi

$$y_i = \alpha + \beta_1(x_{i1} - \overline{x}_1) + \beta_2(x_{i2} - \overline{x}_2) + \cdots + \beta_q(x_{iq} - \overline{x}_q) + \varepsilon_i,$$

Dimana

$$\alpha = \beta_0 + \beta_1\overline{x}_1 + \beta_2\overline{x}_2 + \cdots + \beta_q\overline{x}_q .$$

Untuk menduga β_1, maka :

$$\boldsymbol{\beta}_1 = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_q \end{pmatrix},$$

Dan menggunakan matrik x pusat sbb :

$$\mathbf{X}_c = \begin{pmatrix} x_{11} - \overline{x}_1 & x_{12} - \overline{x}_2 & \cdots & x_{1q} - \overline{x}_q \\ x_{21} - \overline{x}_1 & x_{22} - \overline{x}_2 & \cdots & x_{2q} - \overline{x}_q \\ \vdots & \vdots & & \vdots \\ x_{n1} - \overline{x}_1 & x_{n2} - \overline{x}_2 & \cdots & x_{nq} - \overline{x}_q \end{pmatrix} = \begin{pmatrix} (\mathbf{x}_1 - \overline{\mathbf{x}})' \\ (\mathbf{x}_2 - \overline{\mathbf{x}})' \\ \vdots \\ (\mathbf{x}_n - \overline{\mathbf{x}})' \end{pmatrix},$$

Sehingga penduga β_1 adalah sbb :

$$\hat{\boldsymbol{\beta}}_1 = (\mathbf{X}_c'\mathbf{X}_c)^{-1}\mathbf{X}_c'\mathbf{y}.$$

- **Pengujian kebebasan Antar Variabel Respon**

Variabel $Y_1, \ldots, Y_2, \ldots, Y_q$ dikatakan bersifat saling bebas (independen) jika matriks korelasi antar variabel membentuk matriks identitas. Untuk menguji kebebasan antar variabel ini dapat dilakukan uji Barlett Sphericity berikut (Morrison, 2005) :

Hipotesis :

H_0 : Antar variabel respon bersifat independen

H_1 : Antar variabel respon bersifat dependen

Statistik Uji :

$$X^2_{hitung} = -\left\{n - 1 - \frac{2q+5}{6}\right\} \ln|R|$$

Dimana q adalah jumlah variabel respon dan ln $|R|$adalah nilai-nilai determinan matrik korelasi dari masing-masing variabel respon. Gagal tolak H₀ jika $X^2_{hitung} \leq X^2_{\alpha,\frac{1}{2}q(q+1)}$ yang berarti antar variabel bersifat independen.

- **Kullback's Information Criterion Corrected (KICc)**

Menurut Hafidi dan Mkhadri (2006) criteria KICc merupakan koreksi dari metode KICc dan akan menghasilkan model terbaik jika digunakan pada sampel kecil untuk pemilihan model linear multivariate. Hafidi dan Mkhadri (2006) menyatakan bahwa besarnya KICc adalah

$$KICc = n\left(ln\left|\hat{\Sigma}\right| + q\right) + \frac{d(3n - p - q - 1)}{n - p - q - 1}$$

Dengan :

d = qp + 0,5q(q+1)

q = jumlah variabel respon

n = jumlah data

$\hat{\Sigma}$ = penaksir matriks varian-kovarian error

Kriteria pemilihan model terbaik jika didapatkan nilai KICc terkecil yang berarti semakin kecil nilai dari KICc maka semakin baik model yang digunakan.

- **Hubungan Antara Variabel Respon dan variabel Prediktor**

Pada regresi linear multivariate, ukuran yang digunakan untuk mengukur hubungan antara variabel respond an predictor adalah Wilk's lambda. Ukuran dinyatakan dengan rumus sbb (Rencher, 2002) :

$\eta^2_\Lambda = 1 - \Lambda$ dengan Λ adalah nilai Wilk's lambda

Nilai η_{Λ}^{2} berada pada interval 0 dan 1. Artinya, semakin mendekati 1 berarti hubungan antara variabel respon dan variabel predictor semakin erat. Karena η_{Λ}^{2} melibatkan nilai dari Wilk's lambda yang mana nilai Wilk's lambda memperhitungkan antar variabel respon, hal tersebut dapat ditunjukkan dalam rumus Λ terdapat Y'Y.

- **Uji Hipotesis untuk Model Regresi**

Pengujian ini dilakukan untuk menguji apakah model regresi tidak sama dengan nol. Hipotesis yang digunakan adalah (Rencher, 2001) :

H_0 : $\beta_i=0$ dimana $\beta_i'=(\beta_1, \beta_2, \dots, \beta_q)$

H_1 : $\beta_i \neq 0$.

Statistic uji :

$$F = \frac{\text{SSR}/q}{\text{SSE}/(n-q-1)},$$

Dimana, SSR = $\left(\hat{\beta}' X' y - n\overline{y^2}\right)$ dan SSE = $\left(y' y - \hat{\beta}' X' y\right)$

Tolak H_0 jika $F_{hitung} > F_{a,q,n-q-1}$.

- **Uji Hipotesis untuk Koefisien β**

Pengujian ini dilakukan untuk menguji apakah secara keseluruhan parameter tidak sama dengan nol. Hipotesis yang digunakan adalah sbb (Rencher, 2002) :

$H_0 = \beta_{11} = \beta_{12} = \cdots = \beta_{p1} = \cdots = \beta_{pq} = 0$

H_1: $paling\ sedikit\ ada\ satu\ \beta_{pq} \neq 0$

Dengan B=$\begin{pmatrix} \beta_0^T \\ B_1 \end{pmatrix} = \begin{pmatrix} \beta_{01} & \beta_{02} & \cdots & \beta_{0q} \\ \beta_{11} & \beta_{21} & \cdots & \beta_{1q} \\ \vdots & \vdots & & \vdots \\ \beta_{p1} & \beta_{p2} & \ldots & \beta_{pq} \end{pmatrix}$

Statistik uji yang digunakan adalah Wilk's lambda

$\Lambda = \frac{|E|}{|E+H|} = \frac{|Y'Y - \hat{B}'X'Y|}{|Y'Y - n\bar{y}\bar{y}'|}$ dimana $\bar{y}$ adalah vector rata-rata dari matriks Y

H_0 ditolak jika $\Lambda_{hitung} \leq \Lambda_{a,q,p,n-p-1}$ adalah nilai table kritis untuk Wilk's Lambda.

- **Uji Asumsi IIDN**

Asumsi yang harus dipenuhi dalam melakukan pemodelan regresi multivariate adalah residual memiliki matriks varian-kovarian yang homogen. Untuk menguji syarat ini dapat dipergunakan statistic uji Box's M (Rencher, 2002) sbb :

Hipotesis :

H_0 : $\sum_1 = \sum_2 = \ldots = \sum_j = \sum$

H_1 : minimal ada satu $\sum_i \neq \sum_j$ untuk $i \neq j$

Statistik uji :

H_0 : $\sum_1 = \sum_2 = \ldots = \sum_k = \sum$

H_1 : minimal ada satu $\sum_k \neq \sum_l$ untuk $k \neq l$

Statistic uji :

$$u = -2(1 - C_1) \ln M$$

Dimana :

$$S_{pool} = \frac{\sum_{l=1}^{k} v_k S_k}{\sum_{l=1}^{k} v_k} \qquad \ln M = \frac{1}{2}\sum_{l=1}^{k} v_t \, ln|S_t| - \frac{1}{2}\left(\sum_{l=1}^{k} v_t\right)\ln\left|S_{pool}\right|$$

Dengan $c_1 = \left[\sum_{l=1}^{k} \frac{1}{v_l} - \frac{1}{\sum_{l=1}^{k} v_l}\right]\left[\frac{2p^2+3p-1}{6(p+1)(k-1)}\right] \quad v_l = n_l - 1$

Asumsi residual independen, digunakan uji Barlett Sphercity. Variabel $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_q$ dikatakan bersifat saling bebas (independen) jika matriks korelasi antar residual membentuk matriks identitas. Untuk menguji kebebasan antar residual dapat dilakukan uji Barlett Spericity sbb (Morrison, 2005) :

Hipotesis :

H_0 : residual bersifat independen

H_1 : residual bersifat dependen

Statistic uji :

$$X_{hitung}^2 = -\left\{n - 1 - \frac{2q+5}{6}\right\}\ln|R|$$

Gagal tolak H_0 jika $\chi_{hitung}^2 \leq \chi_{\frac{1}{2}q(q-1)}^2$ yang berarti antar residual bersifat independen

Asumsi yang harus dipenuhi dalam pemodelan regresi linier multivariate adalah residual yang memiliki distribusi multivariate normal. Pemeriksaan distribusi multivariate normal dapat dilakukan dengan cara membuat q-q plot dari nilai d_i^2 (Johnson & Winchern, 2007).

Hipotesis yang digunakan adalah sbb :

H_0 : residual berdistribusi normal multivariate

H_1 : residual tidak berdistribusi normal multivariate

$$d_i^2 = (\hat{\varepsilon}_i - \bar{\varepsilon})'S^{-1}(\hat{\varepsilon}_i - \bar{\varepsilon}), \quad i = 1,2,\ldots,n$$

Kesimpulan adalah gagal tolak H_0 atau data dikatakan berdistribusi normal multivariate jika ada sejumlah data yang memiliki nilai $d_i^2 \leq \chi^2_{q,0,5}$ lebih dari 50%.

- **Korelasi Deterministik terkoreksi (R^2)**

Proporsi total variasi dalam y dapat dibentuk ke dalam regresi pada x yang dinotasikan dengan R^2 sebagai berikut:

$$R^2 = \frac{jumlah\ kuadrat\ regresi}{jumlah\ kuadrat\ total}$$

$$= \frac{\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} - n\bar{\mathbf{y}}^2}{\mathbf{y}'\mathbf{y} - n\bar{\mathbf{y}}^2}.$$

Dimana R^2 disebut dengan koefisien determinasi ganda atau lebih umumnya adalah korelasi berganda. Korelasi berganda R didefinisikan sebagai akar kuadrat postif R^2.

Uji F untuk model regresi dibentuk dari R^2 sebagai :

$$F = \frac{n-q-1}{q}\frac{R^2}{1-R^2}.$$

- **Seleksi Variabel**

Dalam beberapa kasus regresi, banyak ditemukan model yang tidak memasukkan semua variabel yang terlibat di dalamnya. Hal ini diakibatkan oleh adanya validasi model yang nantinya apabila variabel dimasukkan dalam model akan berakibat menurunkan variansi $\hat{\beta_j}$ dan $\hat{y_i}$. Ada beberapa cara untuk menyeleksi variabel-variabel tersebut, yaitu dengan memeriksa semua variabel yang mungkin dan dengan menggunakan teknik stepwise.

Pemeriksaan Semua Variabel

Pemeriksaan ini digunakan untuk mencari variabel yang sesuai dalam model. Dalam beberapa literature menyebutkan jumlah variabel dalam suatu model dinotasikan dengan p-1, sehingga dengan mencantumkan intercept, ada beberapa p parameter didalam model. Sesuai dengan jumlah total variabel yang tersedia dari model yang diseleksi dinotasikan dengan k-1, dengan k parameter dalam model. Ada beberapa cara dalam pemeriksaan ini, yaitu :

1. R_p^2.

   R_p^2 didefinisikan sebagai proporsi total jumlah kuadrat regresi , R^2 mengukur dengan jelas suatu model. p subskrip adalah suatu indeks ukuran subset, dengan menggambarkan jumlah parameter dalam model, termasuk intercept. Bagaimanapun juga, R_p^2 tidak dapat mencapai maksimum untuk beberapa nilai dari $p < k$ karena hal tersebut tidak dapat menurunkan ketika variabel ditambah dalam model.
2. S_p^2.

   Beberapa criteria lain yang berguna adalah penduga variansi untuk setiap subset yang didefinisikan sbb :

$$s_p^2 = \frac{\text{SSE}_p}{n-p}.$$

Untuk p = 2, 3,..., k. Jika k terlalu besar, tipe pola p yang di dekati oleh k adalah untuk meminimumkan S_p^2 untuk pengurangan secara keseluruhan minimum kurang dari S_k^2 dan kemudian meningkatkannya. Nilai minimum S_p^2 dapat kurang dari S_k^2 jika pengurangan SSE_p dengan penambahan variabel yang tidak diimbangi dengan pengurangan derajat bebas. Alternative yang diberikan adalah memilih p seperti meminimumkan p $S_p^2 = S_k^2$ atau memilih nilai yang paling kecil dari p seperti $\min_p S_p^2 < S_k^2$, ketika tidak aka nada p< k seperti $\min_p S_p^2$ persis sama dengan S_k^2.

3. Cp

   Statistic C_p pertama kali dikemukan oleh C.L.Mallows dengan menggunakan rumus sbb :

$$C_p = \frac{\text{SSE}_p}{s_k^2} - (n - 2p).$$

   Dalam hal ini, SSE adalah jumlah kuadrat sisa dari model yang mengandung p parameter, p adalah banyaknya parameter dalam model termasuk β_0 dan S_k^2 adalah kuadrat tengah sisaan dari persamaan terbesar yang dipostulatkan mengandung semua Z, dan diasumsikan merupakan nilai dugaan takbias yang terandalkan bagi galat σ^2. Model terbaik ditentukan setelah memeriksa tebaran C_p. Yang dicari adalah persamaan regresi dengan nilai C_p rendah yang kira-kira sama dengan p. Bila mana yang dipilih tidak jelas, maka merupakan penilaian yang bersifat pribadi untuk memilih :

   1. Persamaan berbias yang tidak diwakili data apapun karena JKS_p nya lebih besar (sehingga $C_p > p$) namun memiliki C_p

yang lebih kecil yang merupakan nilai dugaan bagi simpangan total (galat ragam plus galat bias) dari model sebenarnya tidak diketahui, atau

2. Persamaan dengan parameter lebih banyak sehingga mempunyai ketepatan lebih baik terhadap data (artinya $C_p \approx p$) namun mempunyai simpangan total (galat ragam plus galat bias) yang lebih besar dari model sebenarnya yang tidak diketahui. Dengan kata lain, model yang lebih kecil cenderung mempunyai nilai C_p yang lebih kecil pula, namun model yang lebih besar (dengan p yang lebih besar) cenderung mempunyai C_p yang lebih dekat dengan p.

- **Stepwise Regresi**

Metode eliminasi langkah mundur mulai dengan regresi terbesar dengan menggunakan semua peubah, dan secara bertahap mengurangi banyaknya peubah di dalam persamaan sampai suatu keputusan dicapai untuk menggunakan persamaan yang diperoleh. Prosedur stepwise ini mencapai kesimpulan yang serupa namun dengan menempuh arah yang berlawanan, yaitu menyusupkan peubah satu demi satu sampai diperoleh persamaan regresi yang memuaskan. Urutan penyisipannya ditentukan dengan menggunakan koefisien korelasi parsial sebagai ukuran pentingnya peubah yang masih di luar persamaan. Prosedur dalam penyeleksian menggunakan stepwise ini hamper sama dengan melihat variabel yang terbaik dimasukkan dalam setiap langkah. Kemudian setelah variabel telah dimasukkan, setiap variabel yang baru dimasukkan dengan memeriksa secara individu dengan menggunakan uji F untuk melihat jika variabel tersebut tidak signifikan dan dikeluarkan dari model.

Ada beberapa kemungkinan resiko dalam menggunakan metode stepwise. Dalam prosedur stepwise bisa jadi gagal mendeteksi

variabel predictor yang benar karena S_p^2 adalah bias yang tinggi dibawah spesifikasi model, sehingga mengurangi nilai F pada masing-masing variabel yang dibuat. Di sisi lain, sebuah variabel yang bukan variabel predictor y yang benar dapat memasukkan karena adanya korelasi yang ada dalam sampel tertentu. Dalam beberapa kesempatan, seperti variabel yang kuat, statistic uji F untuk memasukkan variabel yang tidak mempunyai suatu distribusi F karena dimaksimumkan pada setiap langkah.

Langkah-langkah dalam analisis regresi multivariate yaitu sbb :

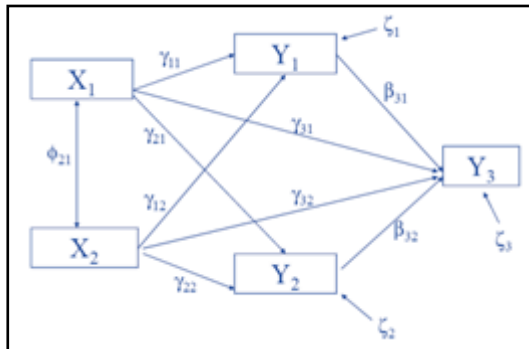
1. Melakukan pengujian korelasi antar variabel respon dengan menggunakan uji Barlett test. Jika terbukti variabel respon berkorelasi maka analisis dapat dilanjutkan pada tahap selanjutnya dengan metode multivariate, jika tidak maka metode yang digunakan adalah metode univariat.
2. Melakukan pemilihan model dengan menggunakan metode KICc berdasarkan factor-faktor yang mempengaruhi.
3. Melakukan pengujian estimasi parameter model regresi multivariate
4. Melakukan pengujian signifikansi parameter model regresi multivariate
5. Melakukan pengujian asumsi residual IIDN (identik, independen, dan distribusi normal)
6. Mendapatkan factor-faktor yang berpengaruh dan kesimpulan.

Dengan demikian, karena adanya korelasi pada regresi multivariate maka akan di cari berapa besar korelasi diantara variabel dependen. Sehingga untuk mengetahui besar korelasi antar variabel dependen maka dapat dilanjutkan kembali dengan menggunakan analisis jalur (path analysis).

14.4. Analisis Jalur (*Path Analysis*)

Analisis jalur merupakan pengembangan lebih lanjut dari analisis regresi berganda ataupun multivariate. Analisis jalur ingin menguji persamaan regresi yang melibatkan beberapa variabel eksogen dan endogen sekaligus sehingga memungkinkan pengujian terhadap variabel mediating/intervening atau variabel antara. Disamping itu, analisis jalur juga dapat mengukur hubungan langsung antar variabel dalam model maupun hubungan tidak langsung antar variabel dalam model. Analisis jalur yang merupakan pengembangan dari model regresi yang digunakan untuk menguji kesesuaian (fit) dari matrik korelasi dari dua atau lebih model yang dibandingkan oleh si peneliti. Model biasanya digambarkan dengan lingkaran dan anak panah yang menunjukkan hubungan kausalitas. Regresi dilakukan untuk setiap variabel dalam model. Nilai regresi yang diprediksi oleh model dibandingkan dengan matrik korelasi hasil observasi variabel dan nilai goodness of fit dihitung. Model terbaik dipilih berdasarkan nilai goodness of fit. Menurut Imam Ghozali, terdapat dua asumsi yang melandasi diagram jalur, yaitu :

1. Semua hubungan kausalitas didasarkan pada teori. Teori sebagai dasar memasukkan atau menghilangkan hubungan kausalitas.
2. Hubungan kausalitas dalam model dianggap linier.

Contoh model analisis jalur:

Dengan notasi-notasi yang digunakan sebagai berikut:

- γ (gamma) : koefisien pengukur hubungan antara variabel endogen dengan eksogen
- β (beta) : koefisien yang mengukur hubungan antar variabel dependen (endogen).
- ϕ (phi) : koefisien yang mengukur hubungan antar variabel independen (eksogen).
- ζ (zeta) : varian peubah latent yg tdk terjelaskan model
- Y : variabel dependen (endogen)
- X : variabel independen (eksogen)

14.5. Ilustrasi Analisis Jalur (Path Analysis)

Misalkan kita ingin melakukan penelitian mengenai model hubungan antara gaji yang diterima oleh pekerja di suatu perusahaan dengan latar belakang pendidikan formal yang dimiliki, masa kerja, gaji awal bekerja, serta masa kerja sebelumnya, serta model hubungan antara gaji awal bekerja dengan latar belakang pendidikan formal serta masa kerja sebelumnya. Diambil 474 data dari pekerja yang ada, dimana:

X1 = pendidikan dalam tahun pendidikan formal

X2 = masa kerja pada pekerjaan (perusahaan) sekarang dalam bulan

X3 = masa kerja sebelum pekerjaan (perusahaan) sekarang dalam bulan

Y1 = Gaji sekarang (dalam Juta Rp)

Y2 = Gaji awal bekerja (dalam Juta Rp)

Misalnya kita membentuk model sebagai berikut:

Y1 = X1   X2   X3   Y2

Y2 = X1   X3

Yang berarti bahwa gaji sekarang (Y1) selain dipengaruhi oleh pendidikan (X1), masa kerja sekarang (X2), masa kerja sebelumnya (X3), juga dipengaruhi oleh gaji awal bekerja (Y2). Gaji awal bekerja (Y2) itu sendiri juga dipengaruhi oleh pendidikan (X1) dan masa kerja sebelumnya (X3).

Input untuk menjalankan persamaan tersebut adalah sebagai berikut:

```
Analis Path Dengan Menggunakan LISREL
Raw Data from file 'D:\blog\kerja1.psf'
Relationships
Y1 = CONSTAN X1 X2 X3 Y2
Y2 = CONSTANT X1 X3
Path Diagram
Options: SS EF
End of Problem
```

Catatan: perintah *Option: SS*, digunakan untuk menampilkan nilai standardized hubungan antar variable. Jika perintah ini tidak digunakan, maka yang akan ditampilkan hanya nilai unstandarized (tetapi nilai standardized tetap bias dilihat pada path diagram). Option: EF digunakan untuk menampilkan effect decomposition (komposisi pengaruh).

Dengan input tersebut, LISREL menghasilkan estimasi regresi unstandarized berikut:

```
Y1 = - 1.61  + 1.77*Y2 + 0.067*X1 + 0.016*X2 - 0.0017*X3, Errorvar.= 0.55  , R² =0.81
      (0.33)  (0.059)   (0.016)    (0.0034)   (0.00035)              (0.036)
      -4.92   30.22     4.06       4.73       -4.91                  15.33

Y2 = - 0.99  + 0.19*X1 + 0.0016*X3, Errorvar.= 0.34  , R² = 0.45
      (0.14)  (0.0097)  (0.00027)             (0.022)
      -6.98   19.40     6.17                  15.33
```

Dari output tersebut terlihat bahwa semua variabel pada taraf 1 % berpengaruh signifikan, baik terhadap Y1 maupun Y2.

Selanjutnya, output covariance matrix untuk independent variables dan latent variables diberikan sebagai berikut. Interpretasinya sama dengan kasus sebelumnya.

```
Covariance Matrix of Independent Variables

               X1         X2         X3
         --------   --------   --------
   X1        8.32
           (0.54)
            15.33

   X2        1.38     101.22
           (1.34)     (6.60)
             1.03      15.33

   X3      -76.14       3.13   10938.28
          (14.35)    (48.54)   (713.53)
            -5.30       0.06      15.33

Covariance Matrix of Latent Variables

               Y1         Y2         X1         X2         X3
         --------   --------   --------   --------   --------
   Y1        2.94
   Y2        1.19       0.62
   X1        3.25       1.44       8.32
   X2        2.19       0.26       1.38     101.22
   X3      -17.41       3.72     -76.14       3.13   10938.28
```

Selanjutnya, sebagian output Goodness of Fit Statistics diberikan dibawah ini:

```
                    Goodness of Fit Statistics
                          Degrees of Freedom = 1
          Minimum Fit Function Chi-Square = 2.41 (P = 0.12)
```

Model memiliki fit yang sangat baik karena memiliki nilai probabilitas yang tidak signifikan (p-value = 0,12 Chi-Square = 2,41 dengan df = 1). Catatan: model yang fit seharusnya memiliki nilai p yang tidak signifikan (lebih besar dari 0,05).

Berdasarkan perintah *Option: SS*, LISREL memberikan output sebagai berikut:

```
Standardized Solution

        BETA

                   Y1          Y2
              --------    --------
      Y1          - -        0.81
      Y2          - -         - -

        GAMMA
                   X1          X2          X3
              --------    --------    --------
      Y1          0.11        0.09       -0.11
      Y2          0.69         - -        0.22
```

Output BETA adalah output LISREL yang berupa matriks hubungan antara sesame variable endogen. Bagian kolom adalah variable endogen independent dan bagian baris adalah variable endogen dependent. Dari output tersebut diketahui nilai standardized pengaruh antara Y2 terhadap Y1 adalah 0,81.

Output GAMMA adalah output LISREL yang berupa matriks pengaruh (standardized) antara variable eksogen (independent) terhadap variable endogen (dependen). Bagian kolom adalah variable eksogen (independent) dan bagian baris adalah variable endogen (dependent).

```
        Correlation Matrix of Y and X
                   Y1          Y2          X1          X2          X3
              --------    --------    --------    --------    --------
      Y1          1.00
      Y2          0.88        1.00
      X1          0.66        0.63        1.00
      X2          0.13        0.03        0.05        1.00
      X3         -0.10        0.05       -0.25        0.00        1.00

        PSI
        Note: This matrix is diagonal.

                   Y1          Y2
              --------    --------
                  0.19        0.55
```

Regression Matrix Y on X (Standardized)

	X1	X2	X3
Y1	0.67	0.09	0.07
Y2	0.69	- -	0.22

yaitu output BETA dan GAMMA.

Selanjutnya, berdasarkan perintah Option: EF, LISREL memberikan output berikut:

Total Effects of X on Y

	X1	X2	X3
Y1	0.40	0.02	0.00
	(0.02)	(0.00)	(0.00)
	18.93	4.73	2.03
Y2	0.19	- -	0.00
	(0.01)		(0.00)
	19.40		6.17

Output diatas memberikan pengaruh total antara variable eksogen terhadap variable endogen. Pengaruh total ini merupakan penjumlahan dari pengaruh langsung (lihat output estimasi regresi unstandarized sebelumnya) dengan pengaruh tidak langsung (lihat output di bawah ini). Misalnya pengaruh total X1 terhadap Y1 sebesar 0,40. Ini adalah penjumlahan dari pengaruh langsung X1 terhadap Y1 sebesar 0,067 dengan pengaruh tidak langsung X1 terhadap Y1 sebesar 0,33.

Pengaruh total X2 terhadap Y1 sama dengan pengaruh langsungnya. Karena hubungan X2 dan Y1 adalah langsung dan tidak memiliki hubungan tidak langsung.

Indirect Effects of X on Y

	X1	X2	X3
Y1	0.33	- -	0.00
	(0.02)		(0.00)
	16.32		6.04
Y2	- -	- -	- -

Output diatas memberikan pengaruh tidak langsung antara variable. Misalnya pengaruh tidak langsung X1 terhadap Y1 adalah sebesar 0,33. Ini diperoleh melalui perkalian antara pengaruh langsung antara X1 terhadap Y2 (variable antara) dengan pengaruh langsung Y2 terhadap Y1. Dengan demikian nilai 0,33 diperoleh melalui 0,19 x 1,77 = 0,33. (Untuk lebih mudah memahaminya, lihat path diagram di bawah).

Total Effects of Y on Y

	Y1	Y2
Y1	- -	1.77 (0.06) 30.22
Y2	- -	- -

Output diatas memberikan pengaruh total variable endogen terhadap variable endogen. Pengaruh total Y2 terhadap Y1 sama dengan pengaruh langsungnya. Karena hubungan antara Y2 dan Y1 adalah langsung dan tidak memiliki hubungan tidak langsung.

Standardized Total and Indirect Effects

Standardized Total Effects of X on Y

	X1	X2	X3
Y1	0.67	0.09	0.07
Y2	0.69	- -	0.22

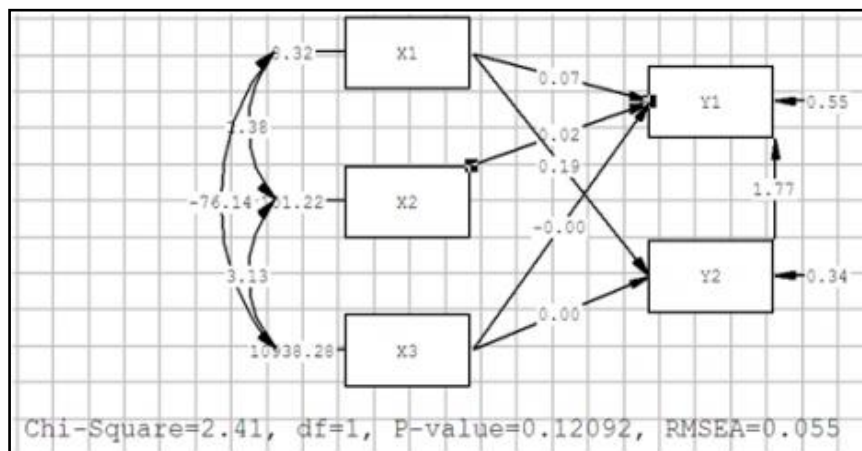
Standardized Indirect Effects of X on Y

	X1	X2	X3
Y1	0.56	- -	0.18
Y2	- -	- -	- -

Standardized Total Effects of Y on Y

	Y1	Y2
Y1	- -	0.81
Y2	- -	- -

Tampilan output diatas sama dengan tampilan sebelumnya, hanya ini merupakan pengaruh yang telah standarized (yang diturunkan dari estimasi nilai standarized).

15

15.Model Persamaan Struktural (*Structural Equation Models*)

15.1. Pendahuluan

Penelitian pada bidang ilmu social umumnya menggunakan konsep-konsep teoritis atau konstruk-konstruk yang tidak dapat diukur atau dapat diamati secara langsung. Namun kita masih bisa menemukan beberapa indicator untuk mempelajari konsep-konsep teoritis tersebut. Kondisi seperti di atas menimbulkan dua permasalahan besar untuk membuat kesimpulan ilmiah dalam ilmu social dan perilaku,yaitu :

1. Masalah pengukuran
2. Masalah hubungan kausal antar variable

Structural Equation Modelling (SEM) merupakan salah satu analisis statistik yang dapat menjawab permasalahan di atas. SEM dikembangkan oleh Karl Joreskorg, Keesling dan Willey. SEM merupakan salah satu tehnik pengolahan data statistik multivariate *structural relationship*. Pemodelan penelitian melalui SEM memungkinkan seorang peneliti dapat menjawab pertanyaan yang regresif maupun dimensional (yaitu mengukur apa dimensi-dimensi sebuah konsep). Pada saat seorang peneliti menghadapi pertanyaan penelitian berupa identifikasi dimensi-dimensi sebuah konsep atau konstruk seperti yang lazim dilakukan pada analisis faktor) dan pada saat yang sama peneliti ingin mengukur pengaruh atau derajat atau hubungan antar faktor yang telah diidentifikasikan dimensi-dimensinya itu,

SEM merupakan alternative jawaban yang layak untuk diajukan. Oleh karena itu, dapat dikatakan bahwa pada dasarnnya SEM adalah kombinasi antara analisis faktor dan analisis regresi berganda.

Terdapat beberapa tujuan dari SEM, antara lain:

1. Mendapatkan model struktural yang dapat digunakan untuk keperluan prediksi. Dalam hal ini, SEM setara dengan REGRESI.
2. Menguji pengaruh (langsung dan tidak langsung) variabel bebas terhadap variabel tidak-bebas, menentukan variabel dominan, dan jalur-jalur keterkaitan antar variabel. Dalam hal ini, SEM setara dengan analisis jalur.

15.2. Konsep Dasar SEM

SEM adalah sekumpulan alat-alat atau teknik statistika yang memungkinkan tidak hanya mendapatkan model hubungan namun juga pengujian sebuah rangkaian hubungan yang relative rumit secara simultan.

Hubungan rumit itu dapat dibangun antara satu atau beberapa variabel dependen dengan satu arah atau beberapa variabel independen. Masing-masing variabel dependen dan independen dapat berbentuk faktor (konstruk yang dibangun dari beberapa variabel indicator). Variabel-variabel itu dapat berbentuk sebuah variabel tunggal yang diobservasi atau yang diukur langsung dalam sebuah penelitian.

Teknik analisis data dalam SEM dilakukan untuk menjelaskan secara menyeluruh hubungan antar variabel yang ada dalam penelitian. SEM digunakan bukan untuk merancang suatu teori, tetapi lebih ditujukan untuk memeriksa dan membenarkan suatu model. Oleh karena itu, syarat utama menggunakan SEM adalah membangun suatu model hipotesis yang terdiri dari model struktural dan model pengukuran dalam bentuk diagram jalur yang berdasarkan justifikasi teori.

15.2.1. Komponen- komponen model SEM

Komponen-komponen dari model SEM dapat dibedakan menjadi:

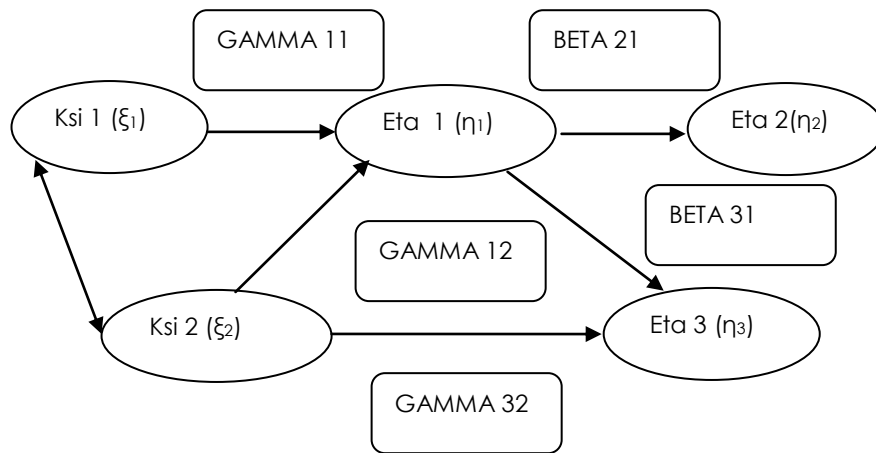
1. 2 jenis variabel laten dan variabel teramati (observed atau measured atau manifest variabel).

   Variabel laten merupakan konsep abstrak sebagai contoh : perilaku orang, sikap, perasaan dan motivasi. Variabel laten ini hanya dapat diamati secara tidak langsung dan tidak sempurna melalui efeknya pada variable teramati. SEM mempunyai 2 jenis variable laten, yaitu :

   - Variabel Eksogen sebagai variabel bebas dan dinotasikan dengan Ksi (ξ).
   - Variabel Endogen merupakan variabel terikat pada model dan dinatasikan dengan Eta (η).

     Variabel laten ini digambarkan dalam bentuk diagram lingkar atau oval atau ellips,

     Variabel teramati adalah variable yang dapat diamati atau dapat diukur secara empiris dan sering disebut sebagai indicator. Variabel terukur adalah variabel yang datanya harus di cari melalui penelitian lapangan dan digambarkan dalam bentuk diagram bujur sangkar seperti

2. 2 jenis model yaitu model struktural (model yang menjelaskan prediksi atau hipotesis hubungan antara variable penyebab terhadap variable akibat)

Gambar 1. model struktural

Keterangan :

ETA1 = GAMMA11 x Ksi 1 + GAMMA12 x Ksi 12

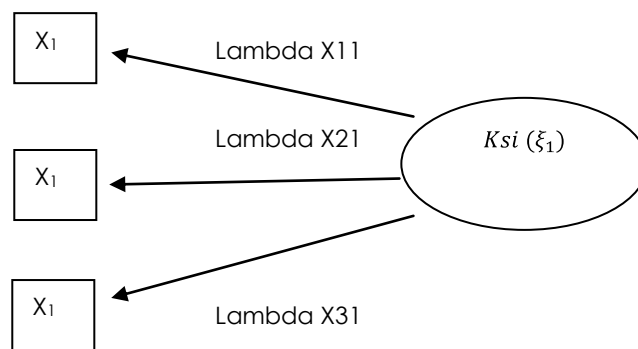
ETA2 = BETA 21 x Eta 1

ETA3 = BETA 31 x Eta 1 + GAMMA 32 x Ksi 12

atau

$$\eta_1 = \gamma_{11}\xi_1 + \gamma_{12}\xi_2$$
$$\eta_2 = \beta_{21}\eta_1$$
$$\eta_3 = \beta_{31}\eta_1 + \gamma_{32}\xi_2$$

dan model pengukuran (model yang menjelaskan operasionalisasi variable penelitian menjadi indikator-indikator terukur yang dinyatakan dalam bentuk diagram jalur dan atau persamaan matematik tertentu). Model pengukuran (*Measurement Model*) adalah proses pemodelan untuk menyelidiki unidimensionalitas dari indicator-indikator yang menjelaskan sebuah faktor. Pengukuran ini bertujuan untuk mengkorfimasi faktor-faktor yang telah terbentuk, sehingga analisis ini disebut *Confirmacy Faktor Analysis*. Keakuratan model

pengukuran adalah dengan pemeriksaan mengenai validitas dan reliabilitas. Hasil yang signifikan dari λ menunjukkan data valid, dan δ maupun ε yang tidak signifikan menunjukkan hasil yang reliable.

Gambar 2. Model Pengukuran

Keterangan :

X1 = Lambda X11 x Ksi 1

X2 = Lambda X21 x Ksi 1

X3 = Lambda X31 x Ksi 1

atau

$$X_1 = \lambda_{x11}\, \xi_1$$
$$X_2 = \lambda_{x21}\, \xi_1$$
$$X_3 = \lambda_{x31}\, \xi_1$$

3. 2 jenis kesalahan yaitu kesalahan struktural dan kesalahan pengukuran.

   Kesalahan struktural muncul disebabkan variabel bebas tidak dapat memprediksi secara sempurna variabel terikat. Kaesalahan ini diasumsikan tidak berkorelasi dengan variable eksogen dari model dan dinotasikan dengan ζ (“zeta”).

Kesalahan pengukuran disebabkan oleh indikator-indikator atau variabel-variabel teramati tidak dapat secara sempurna mengukur variabel laten terkait. Komponen kesalahan yang berkaitan dengan variabel teramati X dinotasikan dengan δ (“delta”), sedangkan yang berkaitan dengan variabel y dinotasikan dengan ε (“epsilon”)

Mengapa SEM dilakukan?

1. Melibatkan variabel laten
2. Evaluasi kualitas pengukuran
3. Memperhitungkan error of measurement dalam manifest variabel (eksogeneous dan endogeneous)
4. Menghasilkan direct indirect dan total effect dari variabel-variabel dalam model

Hubungan antar variabel dinyatakan melalui garis, karena itu bila tidak ada garis berarti tidak ada hubungan langsung yang di hipotesakan.

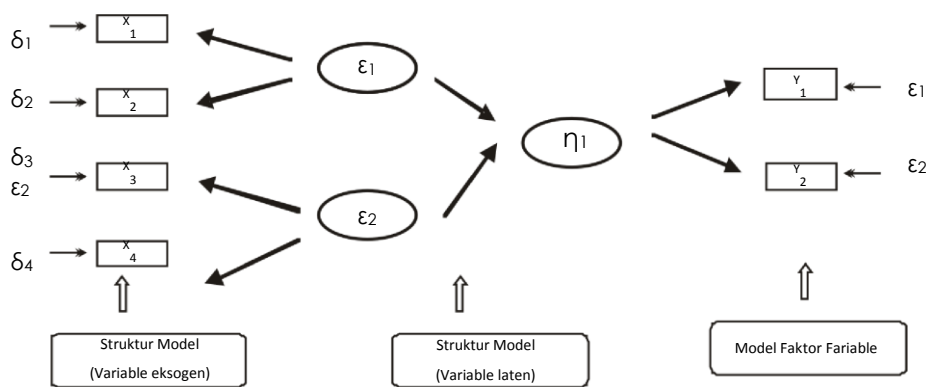
Bentuk-bentuk hubungan antar variabel dalah sebagai berikut :

1. Garis dengan anak panah satu arah (⟶)
   Garis ini menujukkan adanya hubungan yang dihipotesakan antara dua variabel dimana variabel yang dituju oleh anak panah merupakan variabel dependent. dalam SEM terdapat 2 hipotesis dengan anaka panah satu arah, yaitu :
   a. Hipotesa mengenai dimensi faktor
      Dimensi-dimensi sebuah faktor akan terlihat dalam diagram SEM melalui arah anak panah (⟶) yang digunakan. Masing-masing indicator sebagai variabel dependen secara bersama-sama dihipotesakan sebagai dimensi dari sebuah konsep atau faktor.

b. Hipotesa mengenai hubungan regresi

Hipotesa mengenai pengaruh satu atau beberapa varibel independen terhadap satu atau beberapa variabel dependen dinyatakan pula dalam anak panah satu arah (⟶)

2. Garis dengan anak panah dua arah (⟷)

Anak panah dua arah ini digunakan untuk menggambarkan kovarian atau korelasi antara dua variabel

Gambar 3. model Diagram lintasan

Gambar di atas menunjukkan bahwa model persamaan structural merupakan pendekatan terintegrasi antara analisis faktor konfirmatori, model struktural dan analisis jalur.

15.2.2. Notasi SEM

Notasi-notasi yang dipakai dalam SEM :

X = indicator variabel eksogen

Y = indicator varibel endogen

ξ = Ksi, variabel laten X atau variabel laten eksogen

η = Eta, variabel laten Y atau variabel laten endogen

δ = Delta, galat pengukuran pada variabel manifest untuk variabel laten X

ε = Epsilon, Galat pengukuran pada variabel manifest untuk variabel laten Y

γ = Gamma, koofisien pengaruh variabel eksogen terhadap variabel endogen

β = Beta, koofisien pengaruh variabel endogen terhadap variabel eksogen

λ = Lambda, loading faktor

Dalam SEM dapat dilakukan analisis hubungan antar peubah laten dengan peubah indicator dengan metode analisis Faktor Konfirmatori, sekaligus mendapatkan model yang bermanfaat untuk prediksi atau prakiraan, dilakukan dengan model struktural.

Terdapat beberapa model yang dianalisis dalam SEM yaitu :

1. **Model Analisis Faktor**

   Model ini menganalisis faktor-faktor yang mempengaruhi atau menjadi indicator bagi variable laten maupun terikat. model pengukuran yang digunakan dalam Model pengukuran yang digunakan dalam hal ini disebut sebagai CFA model (*Confirmatory Factor Anaysis Model)*. Model pengukuran ini sedikit berbeda dengan model pengukuran sebelumnya, yaitu EFA model *(Exploratory Factor Analysis)*.

   Pada EFA, model rinci yang menunjukkan hubungan antara variabel laten dengan variabel teramati tidak dispesifikasikan dan ditentukan terlebih dahulu (sebelum analisis dilakukan), semua variabel laten diasumsikan mempengaruhi semua variabel teramati dengan error yang tidak boleh berkorelasi. Sedangkan pada CFA model, jumlah variabel dan pengaruh suatu variabel laten terhadap variabel teramati ditentukan terlebih dahulu, kesalahan pengukuran bisa berkorelasi dan dibutuhkan identifikasi parameter. Selain itu beberapa

efek langsung variabel laten terhadap variabel teramati dapat ditetapkan sama dengan nol atau suatu konstanta, begitu pula dengan kovarian variabel–variabel laten dapat diestimasi (ditetapkan) pada nilai tertentu.

Dari EFA bisa ditemukan struktur data tertentu. Sementara dari CFA dapat dinilai kecocokan data dengan model untuk menentukan penerimaan atau penolakan suatu teori.

2. **Model Analisis Regresi**, model ini akan menganalisis antara variabel laten X sebagai variabel laten bebas terhadap variabel laten Y (variabel laten terikat) untuk mendapatkan model struktural yang dapat digunakan saat prediksi.
3. **Model Analisis Jalur (*Path Analysis*)**, model ini akan merepresentasikan sistem persamaan simultan melalui diagram jalur, penguraian ragam dan korelasi dalam model parameter dan pemisahan antara pengaruh langsung, pengaruh tidak langsung serta pengaruh total dari suatu variabel terhadap variabel lain.

15.2.3. Metode Perhitungan

Model persamaan SEM sangat komplek sehingga tidak dihitung secara manual. Model ini menggunakan input data korelasi dan kovarian yang dirumuskan dengan:

Kovarian untuk data populasi,

$$\sigma_x^2 = \frac{\sum(X - \mu_x)^2}{N_{pop}} \qquad \sigma_{xy}^2 = \frac{\sum(X - \mu_x)\ \ (Y - \mu_y)}{N_{pop}}$$

dan untuk data sample,

$$s_x^2 = \frac{\sum(X - \bar{X})^2}{n - 1} \qquad s_{xy}^2 = \frac{\sum(X - \bar{X})\ \ (Y - \bar{Y})}{n - 1}$$

$$s_y^2 = \frac{\sum(Y - \bar{Y})^2}{n-1}$$

Sedangkan rumus korelasi untuk data populasi adalah,

$$\rho_{yx} = \frac{\sigma_{xy}}{\sigma_x \sigma_y} \qquad Y = \alpha + \beta_{yx} X + \varepsilon$$

$$\beta_{xy} = \frac{\sigma_{xy}}{\sigma_x^2} = \rho_{yx} \frac{\sigma_y}{\sigma_x}$$

dan korelasi data sample,

$$r_{yx} = \frac{\sigma_{xy}}{\sigma_x \sigma_y} \qquad Y = \alpha + \beta_{yx} X + \varepsilon$$

$$b_{xy} = \frac{s_{xy}}{s_x^2} = r_{yx} \frac{s_y}{s_x}$$

Input data korelasi dan kovarians yang akan digunakan bisa ditulis dalam bentuk matriks menjadi,

$$P = \begin{bmatrix} 1 & \rho_{12} & \cdots \\ \rho_{21} & 1 & \cdots \\ \rho_{31} & \rho_{32} & \cdots \end{bmatrix} \underset{dibagi\ \sigma_i\sigma_j}{\overset{dikali\ \sigma_i\sigma_j}{\rightleftarrows}} \Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_2^2 & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_3^2 \end{bmatrix}$$

dan untuk data sample menjadi,

$$R = \begin{bmatrix} 1 & r_{12} & r_{13} \\ r_{21} & 1 & r_{23} \\ r_{31} & r_{32} & 1 \end{bmatrix} \underset{dibagi\ \sigma_i\sigma_j}{\overset{dikali\ \sigma_i\sigma_j}{\rightleftarrows}} \Sigma = \begin{bmatrix} s_1^2 & s_{12} & s_{13} \\ s_{21} & s_2^2 & s_{23} \\ s_{31} & s_{32} & s_3^2 \end{bmatrix}$$

Demikian juga dengan model pengukuran bisa kita tuliskan dalam bentuk matrik sebagai berikut:

$$\begin{bmatrix} X_1 \\ X_2 \\ \cdot \\ \cdot \\ X_6 \end{bmatrix} = \begin{bmatrix} \lambda x_{11} & 0 \\ \lambda x_{21} & 0 \\ \cdot & \cdot \\ \cdot & \cdot \\ 0 & \lambda x_{62} \end{bmatrix} \begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix} + \begin{bmatrix} \delta_1 \\ \delta_2 \\ \cdot \\ \cdot \\ \delta_6 \end{bmatrix} \quad dan \quad \begin{bmatrix} Y_1 \\ Y_2 \\ \cdot \\ \cdot \\ Y_6 \end{bmatrix} = \begin{bmatrix} \lambda y_{11} & 0 \\ \lambda y_{21} & 0 \\ \cdot & \cdot \\ \cdot & \cdot \\ 0 & \lambda y_{62} \end{bmatrix} \begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \cdot \\ \cdot \\ \varepsilon_6 \end{bmatrix}$$

atau menjadi,

$X = \Lambda_x\xi + \delta$ dan $Y = \Lambda_y\eta + \varepsilon$

dengan Λ_i adalah matrik model pengukuran.

Dan untuk matrik model struktural menjadi,

$$\begin{bmatrix}\eta_1\\ \eta_2\end{bmatrix} = \begin{bmatrix}0 & 0\\ \beta_{21} & 0\end{bmatrix}\begin{bmatrix}\eta_1\\ \eta_2\end{bmatrix} + \begin{bmatrix}\gamma_{11} & \gamma_{12}\\ \gamma_{21} & \gamma_{22}\end{bmatrix}\begin{bmatrix}\xi_1\\ \xi_2\end{bmatrix} + \begin{bmatrix}\zeta_1\\ \zeta_2\end{bmatrix} \text{ atau } \eta = B\eta + \Gamma\xi + \zeta$$

15.2.4. Asumsi-asumsi SEM :

a. **Ukuran sampel**.

   Ukuran sampel yang harus dipenuhi dalam pemodelan ini adalah minimum berjumlah 100 dan selanjutnya menggunakan perbandingan lima observasi untuk setiap *estimated parameter.*

   Penentuan besarnya sampel:

   1. Kalau pendugaan dg metode Maximum likelihood,maka sampel 100-200, minimum absolut 50
   2. Jumlah sampel = 5-10 kali banyaknya parameter
   3. Jumlah sampel = 5-10 kali jumlah variabel manifest dari keseluruhan variabel laten.

b. Normalitas dan Linearitas

   Normalitas dapat diuji dengan melihat gambar histogram atau dengan metode-metode statistik. uji linearitas dapat dilakukan dengan mengamati scatter plot dari data, yaitu dengan memilih pasangan data dan dilihat pola penyebarannya untuk menduga adanya linearitas. Sebaran data yang dianalisis harus memenuhi asumsi sebaran normal dan hubungan antar estimated parameter bersifat linear.

c. **Outliers**

Outliers adalah observasi yang muncul dengan nilai-nilai ekstrim baik secara univariat maupun multivariate yaitu yang muncul karena kombinasi karakteristik unik yang dimilikinya dan terlihat sangat jauh berbeda dari observasi-observasi lainnya, dan ini bisa mengganggu pada saat analisis data.

d. **Multikolinearitas dan singularitas**

variabel yang saling berhubungan akan menyebabkan hasil yang bias sebaiknya data tidak ada multikolinearitas dan singularitas. Bila ada sebaiknya data dikeluarkan, atau alternative lain adalah data tersebut dibuat *'composit variabels',* dan variabel komposit dapat di analisis lebih lanjut. Multikolinearitas dapat di deteksi dari determinan matriks kovarian. Nilai determinan matriks kovarian yang sangat kecil memberikan indikasi adanya problem multikolinearitas.

15.2.5. Langkah-langkah SEM

Sebuah pemodelan SEM yang lengkap pada dasarnya terdiri dari Measurement berdasarkan indicator-indikator empirisnya.Model dan Structural Model. Model pengukuran ditujukan untuk mengkorfimasi sebuah dimensi atau faktor berdasarkan indicator-indikator empirisnya. Struktural Model adalah model mengenai struktur hubungan yang membentuk atau menjelaskan kausalitas antar faktor.

1. **Pengembangan Model Konsep dan Teori**

   Model Hipotetik = Model Konseptual = Model Teoritis

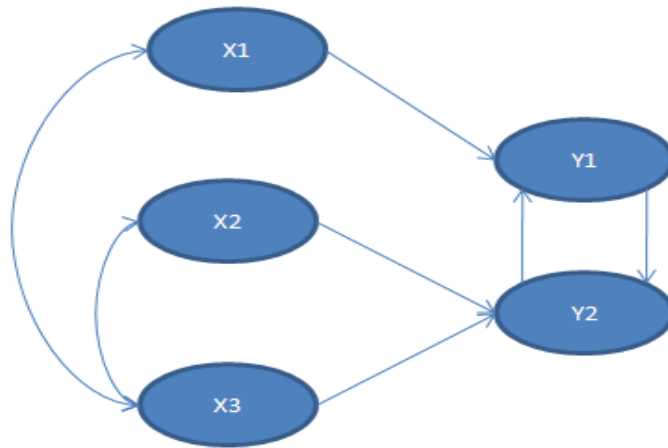
   Hubungan kausalitas sebab-akibat antara variabel eksogen (variabel bebas, independent) dan variabel endogen (variabel tergantung, variabel dependent). Dengan demikian landasan teorinya harus kuat

untuk dapat menjelaskan Model Hipotetik tersebut. Salah satu aspek kritis dalam hal ini adalah "Spesifikasi variabel", terutama variabel prediktif. Untuk kepentingan praktis analisis data dan interpretasinya, maka seyogyanya banyaknya variabel tidak lebih dari 20.

2. Konstruksi Diagram Lintasan

Diagram ini sangat bermanfaat untuk menunjukkan alur-alur (lintasan) kausalita antar variabel yang secara teoritis layak

Hubungan kausalita : Simbol panah satu arah

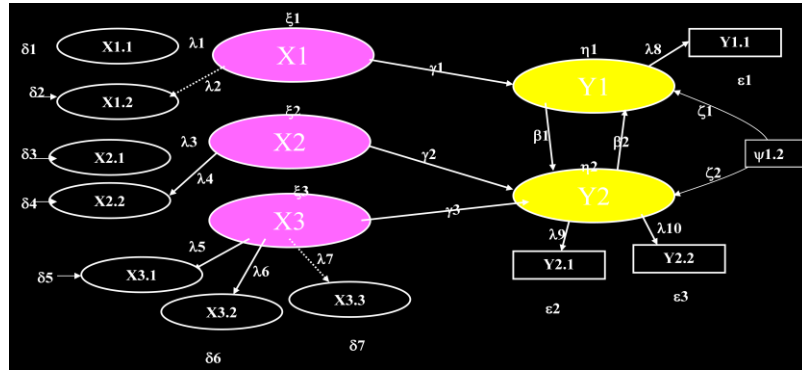
Hubungan korelasional : Simbol panah bolak-balik

Gambar 4. Diagram Lintasan

3. Konversi Diagram Lintasan menjadi Model Struktural

$$\eta 1 = \beta 2.\ \eta 2 + \gamma 1.\ \xi 1 + \zeta 1$$

$$\eta 2 = \beta 1.\ \eta 1 + \gamma 2.\ \xi 2 + \gamma 3.\ \xi 3 + \zeta 2$$

$$X1.1 = \lambda 1.\ \xi 1 + \delta 1$$

$$X1.2 = \lambda 2.\ \xi 1 + \delta 2 \quad \text{.... Dst.}$$

Gambar 5. Ilustrasi path model SEM

4. Memiliki Matriks Input

Input data untuk SEM dapat berupa:
1. Matriks korelasi, atau
2. Matriks peragam, kovarians

Matriks korelasi, digunakan kalau:
1. Tujuannya ingin membuktikan hubungan kausalitas antar variabel
2. Lintasan mana saja yang pengaruhnya dominan
3. Variabel eksogen mana saja yang pengaruhnya dominan terhadap
variabel endogen

Matriks Peragam, digunakan kalau:
1. Tujuannya menguji model hipotetik yang secara teoritis sudah layak
2. Serupa dengan analisis regresi
3. Model yang diperoleh dapat digunakan untuk prediksi
4. Model yg diperoleh dapat untuk menjelaskan fenomena yang dikaji

5. Menilai Masalah Identifikasi

Problematik pendugaan parameter:
1. Un-identified atau under identified
2. Over identified

Gejala yg muncul akibat dari adanya "masalah identifikasi":
a. Adanya standard error yang terlalu besar
b. Matriks informasi tidak dapat disajikan sebagaimana mestinya

c. Nilai penduga parameter tidak dapat diperoleh
d. Muncul angka (nilai) yang aneh
e. Adanya koefisien korelasi yg tinggi (> 0.9) antar koefisien hasil pendugaan

Cara mengatasinya:

1. Landasan teori yang digunakan untuk menyusun Model Hipotetik harus benar-benar 'bagus"
2. Menambah atau mengurangi variabel laten, disesuaikan dengan landasan teorinya
3. Iterasi dalam pendugaan model dengan menetapkan "kendala" pada model, misalnya salah satu atau beberapa parameter model dianggap "fixed"

6. Evaluasi Goodness of fit

Asumsi-asumsi SEM:

1. Asumsi yang berkaitan dengan model
2. Asumsi yang berkaitan dengan pendugaan parameter & pengujian hipotesis

Asumsi yang berkaitan dengan model:

1. Semua hubungan berbentuk linier (Lihat diagram pencarnya)
2. Model bersifat aditif, sesuai dengan landasan teorinya

Asumsi Pendugaan parameter & Uji hipotesis:

1. Random sampling
2. Tidak boleh ada missing data
3. Tidak ada data pencilan, outliers
4. Untuk pendugaan parameter, jumlah sampel minimum 100

Tahapan Goodness of fit :

1. Uji Parameter, dengan t-test:
   a. Parameter Lamda: validitas instrument
   b. Parameter Delta dan Epsilon: Reliabilitas instrumen
   c. Parameter Beta dan Gama, dan lainnya

2. Uji Keseluruhan Model
   Model ini merupakan integrasi antara model struktural dan model pengukuran

3. Uji Model Struktural
   Menggunakan uji koefisien determinasi, seperti model regresi

4. Uji Model Pengukuran
   Uji validitas ……… koefisien korelasi
   Uji reliabilitas ……….. Nilai error

7. Interpretasi dan Modifikasi Model

Bilamana model yang diperoleh cukup baik, maka selanjutnya adalah melakukan interpretasi. Bila model belum baik maka perlu diadakan modifikasi. Modikasi model hanya dapat dilakukan bila ada justifikasi teoritis atau konsep yang cukup kuat sebab metode SEM bukan ditujukan untuk menghasilkan teori atau model, tetapi menguji model. Kelemahannya terutama adalah sangat sulitnya diperoleh diperoleh diperoleh model yang cocok dengan data (fitting model) karena kompleksnya hubungan.

Interpretasi Model:

1. Model Struktural: Interpretasi terhadap fenomena yg sedang dikaji, dan melakukan prediksi
2. Analisis Lintasan:
   a. Efek langsung
   b. Efek tidak langsung
   c. Total efek
   d. Faktor dominan
   e. Kausalitas antar variabel.

15.2.6. Uji Kesesuaian dan Uji Statistik

Dalam analisi SEM tidak ada alat uji statistik tunggal untuk mengukur atau menguji hipotesis mengenai model. Evaluasi goodness of fit yang dimaksud adalah untuk mengukur kebenaran model yang di ajukan. Berikut ini dalah beberapa indek kesesuaian dan cut off value nya yang digunakan untuk menguji apakahsebuah model dapat diterima atau ditolak.

a. Chi-square Statistics

alat uji paling fundamental untuk mengukur overall fit adalah Chi-Square. chi-square ini bersifat sangat sensitive terhadap besarnya sampel yang digunakan karena itu bilamana jumlah smapel yang digunakan cukup besar yaitu > 2000 sampel, chi-square harus didampingi oleh alat uji lain, maka sampel yang digunakan antara range 100-200 sampel. Semakin kecil nilai Chi-square maka semakin kecil kebenaran model tersebut.

b. RMSEA(The Root Means Square Error of Approximation)

RMSEA adalah sebuah indeks yang dapat digunakan untuk mengkompensasi chi-square statistics dalam sampel besar. Nilai RMSEA lebih kecil atau sama dengan 0.08 merupakan indeks untuk diterimanya model yang menunjukkan sebuah close fit dari sebuah model berdasarkan derajat bebas.

c. GFI (Goodness-of fit Index)

Indeks kesesuaian ini akan menghitung proporsi tertimbang dari varian dari matriks kovarian sampel yang dijelaskan oleh matriks kovarian populasi yang tersetimasikan. GFI adalah ukuran non-statistical yang mempunyai rentang nilai antara 0(poor fit) sampai dengan 1(perfect fit). Nilai yang tinggi dalam indeks ini menunjukkan sebuah better fit.

d. AGFI(Adjusted Goodness of fit Index)

Tingkat penerimaan yang direkomendasikan adalah bila AGFI mempunyai nilai yang sama dengan atau lebih besar dari 0.9.

e. CMIN/DF

The minimum Sample Discrepancy Function(CMIN) dibagi dengan derajat bebas menghasilkan indeks CMIN/DF, yaitu salah satu indicator untuk mengukur tingkat kesesuaian sebuah model. Nilai chi-square yang diharapkan adalah kurang dari atau sama dengan 2.00

f. TLI (Tucker Lewis Index)

TLI adalah sebuah alternative incremental fit index yang membandingkan sebuah model yang diuji terhadap sebuah baseline model. Nilai yang direkomendasikan sebagai acuan untuk diterimanya sebuah model adalah penerimaan lebih besar atau sama dengan 0.95.

g. CFI (Comparative Fit Index)

Nilai CFI yang direkomendasikan adalah $\geq$ 0.95. Semakin mendekati 1, maka model semakin baik. Keunggulan dari indeks ini adalah bahwa besaran ini besarnya tidak dipengaruhi oleh ukuran sampel.

Indeks-indeks yang dapat digunakan untuk menguji kelayakan sebuah model, seperti yang tertera di atas, dapat diringkas dalam tabel berikut :

Goodness of fit	Cut-off	Keterangan
Chi-square	Non-signifikan	Dipakai untuk n = 100-200, Model yg baik bila nilai Chi-square tidak jauh berbeda dg derajat bebasnya
RMR	Kecil	Dipakai untuk n besar
RMSEA	< 0.08	Dipakai untuk n besar
GFI	> 0.90	Mirip dengan koef. determinasi R^2
AGFI	>0.9	Mirip dgn R^2 adjusted
CFI	>0.94	Tdk sensitif thd besar sampel
AIC	Kecil	Bila model lebihdari satu, pilihlah yg nilainya kecil

15.3. Penerapan SEM

SEM banyak digunakan pada penelitian-penelitian yang variabel-variabelnya tidak dapat diukur secara langsung. Penelitian-penelitian ini banyak ditemukan di bidang social (ekonomi, sosiometrik, manajemen dan lainnya). dalam perkembangan penggunaan metode SEM didorong

oleh perkembangan berbagai software SEM seperti LISREL, AMOS, ROMANO, SEPATH, dan LISCOMP. Namun demikian diantara software tersebut, LISREL paling banyak digunakan karena selain kemampuan lisrel dalam mengestimasi berbagai masalah SEM, tampilan LISREL juga paling informative dalam menyajikan hasil-hasil statistic. Berikut merupakan contoh penerapan metode SEM dengan menggunakan software LISREL.

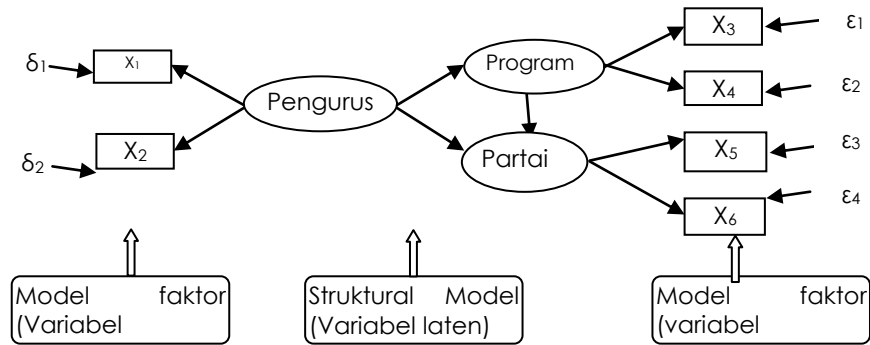
Suatu penelitian mengenai keyakinan masyarakat terhadap partai politik di Indonesia. untuk itu dambil sampel sebanyak 932.

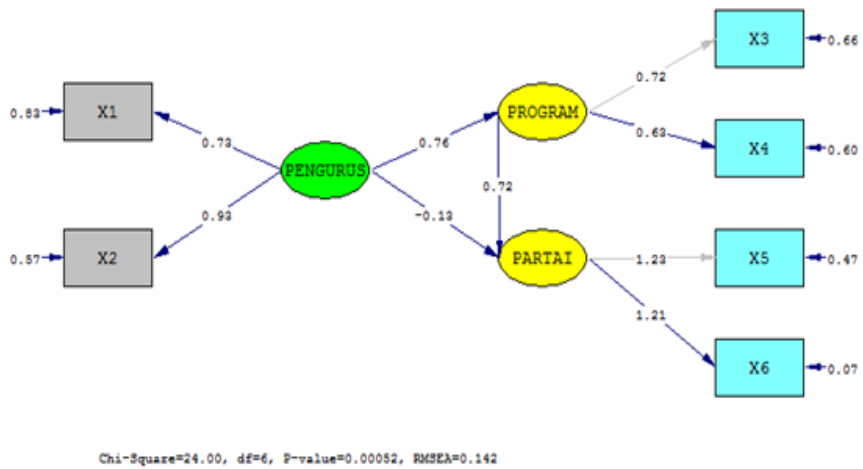
Kerangka teori :

1. keyakinan terhadap partai politik (PARTAI) ditentukan oleh keyakinannya terhadap keberhasilan program Partai (PROGRAM)) dan keyakinan terhadap kemampuan kemampuan pengurus pengurus partai (PENGURUS)
2. Keyakinan terhadap kemampuan pengurus partai juga mempengaruhi keyakinan terhadap keberhasilan program partai

Selanjutnya untuk mengukur keyakinan terhadap partai politik digunakan indikator (X5 dan X6), untuk mengukur keyakinan terhadap keberhasilan program partai digunakan dua indikator (X3 dan X4) sedangkan untuk mengukur keyakinan terhadap kemampuan pengurus partai dengan indicator (X1 dan X2). Rincian indicator dan variabel dapat dilihat pada tabel berikut :

Variabel Laten (Konstruk)	**Variabel Manifest (Terukur)**
Pengurus	X1
	X2
Program partai	X3
	X4
Partai Politik	X5
	X6

Model grafis dari kerangka teoritis diatas adalah :

Gambar 6. Model grafis kerangka teoritis

Dengan menggunakan program LISREL akan didapatkan output dalam bentuk teks dan output grafik.

Gambar 7. Diagram Lintasan

Covariance Matrix

	X3	X4	X5	X6	X1	X2
X3	1.19					
X4	0.45	1.00				
X5	0.57	0.39	1.98			
X6	0.62	0.41	1.49	1.54		
X1	0.34	0.47	0.44	0.22	1.36	
X2	0.46	0.46	0.61	0.52	0.67	1.42

Number of Iterations = 12

LISREL Estimates (Maximum Likelihood)

Measurement Equations

```
X3 = 0.72*PROGRAM, Errorvar.= 0.66 , R² = 0.44
                      (0.12)
                       5.76

X4 = 0.63*PROGRAM, Errorvar.= 0.60 , R² = 0.39
     (0.11)           (0.096)
      5.51             6.32

X5 = 1.23*PARTAI, Errorvar.= 0.47 , R² = 0.76
                    (0.14)
                     3.31

X6 = 1.21*PARTAI, Errorvar.= 0.074, R² = 0.95
     (0.12)          (0.13)
     10.07            0.58

X1 = 0.73*PENGURUS, Errorvar.= 0.83 , R² = 0.39
     (0.11)           (0.14)
      6.70             6.12

X2 = 0.93*PENGURUS, Errorvar.= 0.57 , R² = 0.60
     (0.12)           (0.17)
      7.88             3.34
```

Structural Equations

```
PROGRAM = 0.76*PENGURUS, Errorvar.= 0.42 , R² = 0.58
          (0.14)           (0.20)
           5.58             2.08

PARTAI = 0.72*PROGRAM - 0.13*PENGURUS, Errorvar.= 0.60 , R² = 0.40
         (0.30)         (0.26)             (0.14)
          2.36          -0.49               4.37
```

Reduced Form Equations

PROGRAM = 0.76*PENGURUS, Errorvar.= 0.42, R^2 = 0.58
(0.14)
5.58

PARTAI = 0.42*PENGURUS, Errorvar.= 0.82, R^2 = 0.18
(0.10)
4.01

Correlation Matrix of Independent Variables

PENGURUS
1.00

Covariance Matrix of Latent Variables

	PROGRAM	PARTAI	PENGURUS
PROGRAM	1.00		
PARTAI	0.62	1.00	
PENGURUS	0.76	0.42	1.00

Goodness of Fit Statistics
Degrees of Freedom = 6
Minimum Fit Function Chi-Square = 24.80 (P = 0.00037)
Normal Theory Weighted Least Squares Chi-Square = 24.00 (P = 0.00052)
Estimated Non-centrality Parameter (NCP) = 18.00
90 Percent Confidence Interval for NCP = (6.53 ; 37.02)

Minimum Fit Function Value = 0.17
Population Discrepancy Function Value (F0) = 0.12
90 Percent Confidence Interval for F0 = (0.044 ; 0.25)
Root Mean Square Error of Approximation (RMSEA) = 0.14
90 Percent Confidence Interval for RMSEA = (0.085 ; 0.20)
P-Value for Test of Close Fit (RMSEA < 0.05) = 0.0057

Expected Cross-Validation Index (ECVI) = 0.36
90 Percent Confidence Interval for ECVI = (0.29 ; 0.49)
ECVI for Saturated Model = 0.28
ECVI for Independence Model = 2.61
Chi-Square for Independence Model with 15 Degrees of Freedom = 376.22
Independence AIC = 388.22
Model AIC = 54.00
Saturated AIC = 42.00
Independence CAIC = 412.28
Model CAIC = 114.16
Saturated CAIC = 126.22

1. Analisis terhadap Hasil Estimasi.

   Pada hasil out put diatas, tidak ada error variances yang negative, semua varibel mempunyai t-values > 1,96, berarti estimasi muatan faktor tersebut signifikan. dan kecuali X4 (0.63) semua nilai muatan faktor standar >0,7. Dengan demikian dapat disimpulkan bahwa hasil estimasi muatan faktor dari model baik.

2. Uji kecocokan model

   Uji kecocokan keseluruhan model

   Nilai Chi-square = 24,00 dan p-value = 0,00052 < 0,05. dapat disimpulkan bahwa dari chi-square kecocokan model sudah baik. NCP (Non Centrality Parameter)=18,00 yang merupakan nilai yang cukup kecil dan 90% confident interval dari NCP sehingga berdasarkan NCP dapat disimpulkan kecocokan keseluruhan model baik. RMSEA = 0,14 > 0,08 yang menunjukkan keseluruhan model kurang baik dan p-value (RMSEA < 0,05) = 0,0057 < 0,005, kecocokan model kurang baik. Berikut ini adalah tabel hasil rangkuman analisis terhadap ukuran-ukuran *goodness Of Fit* (GOF):

Ukuran GOF	**Target-tingkat**	**Hasil estimasi**	**Tingkat Kecocokan**
Chi-square P	Nilai yang kecil P > 0,05	$\chi^2 = 24$ (p=0.000)	Baik
NCP Interval	Nilai yang kecil	18,00 (6.53 ; 37,02)	Baik
RMSEA	RMSEA ≤ 0,08	0,14	Kurang baik

P (close fit)	P ≥ 0,5	P = 0,0057	
ECVI	Nilai yang kecil dan dekat dengan ECVI saturated	M* = 0,36 S* = 0,28 I* = 2,61	Kurang baik
AIC	Nilai yang kecil dan dekat dengan AIC saturated	M* = 54,00 S* = 42,00 I* = 388,22	Baik
CAIC	Nilai yang kecil dan dekat dengan CAIC saturated	M* = 114,16 S* = 126,22 I* = 412,28	Baik
NFI	NFI ≥ 0,90	0,93	Baik
NNFI	NNFI ≥ 0,90	0,87	Baik
GFI	GFI ≥ 0,90	0,95	Baik
IFI	IFI ≥ 0,90	0,95	Baik
RFI	RFI ≥ 0,90	0,84	Baik
CN	CN ≥ 200	102,03	Kurang baik
RMR	Standardized RMR ≤ 0,05	0,047	Baik
GFI	GFI ≥ 0,90	0,95	Baik
AGFI	AGFI ≥ 0,90	0,82	Kurang baik

Dari tabel dapat dilihat bahwa ada 4 ukuran GOF yang menunjukkan kecocokan yang kurang baik dan 11 ukuran GOF yang menunjukkan kecocokan yang baik, sehingga kita dapat menyimpulkan bahwa kecocokan keseluruhan model adalah baik. Meskipn demikian, kecocokan keseluruhan model ini masih bisa ditingkatkan dengan memanfaatkan informasi yang ada di *Modification index*.

Analisis Model Pengukuran

Evaluasi terhadap validitas (*validity*) dari model pengukuran dapat dilihat dari nilai t muatan faktor (*factor loading*) pada variabel teramati terhadap laten dari model. Suatu variabel mempunyai validitas yang baik terhadap konstruk atau variabel latennya jika :

- Nilai muatan faktornya lebih besar dari kritis (≥1,96 atau untuk praktisnya ≥2)
- Muatan faktor standarnya (*standardized factor loadings*) ≥ 0,70 (Rigdon dan Ferguson, 1991) atau ≥ 0,50 (Igbaria.et.al., 1997).

Dari output dan gambar 7 diperoleh semua nilai t muatan faktor > 2, jadi faktor-faktor dari variabel-variabel yang ada dalam model adalah signifikan. Kecuali muatan faktor standar (SLF) dari X4= 0,62 (> 0,50), semua faktor standar lainnya > 0,70. Dengan demikian dapat disimpulkan bahwa validitas semua variabel teramati terhadap variabel latennya adalah baik.

Evaluasi terhadap reliabilitas (reliability) dari model pengukuran dapat dilakukan dengan model composite realibity measure (ukuran reliabilitas komposit) dan *variance extracted measure* (ukuran ekstrak varian). Hasil perhitungan reliabitas diperoleh semua nilai Construct reliability (CR) ≥0,70, dan semua nilai Variance Extracted (VE) ≥ 0,50. Dengan demikian dapat disimpulkan bahwa reliabilitas model pengukuran (konstruk) adalah baik

Analisis model struktural

t-value dari koofisien/ parameter :

- Pengurus → program : 5,58 > 2, koofisien signifikan
- Program → partai : 2,08 >2, koofisien signifikan
- Pengurus → partai : -0,49; absolut (-0,49) < 2 koofisien tidak signifikan

Nilai koofisien/ parameter :

- Pengurus → program : 0,76
- program → partai : 0,72
- pengurus → partai : -0,23

Koofisien determinasi (R^2)

Menurut Joreskog (R^2) pada Structur equation tidak mempunyai interpretasi yang jelas dan untuk menginterpretasikan R^2 seperti pada persamaan regresi, harus menggunakan reduced form equation :

- pengurus → Program : 0,58; berarti 58% dari variasi pada partai dijelaskan oleh variasi Pengurus.
- Pengurus → Partai : 4,01, berarti 40,1% dari variasi pada partai dijelaskan oleh variasi dari pengurus.

Hasil dari evaluasi ini bisa dirangkum pada tabel dibawah ini ditambah dengfan asumsi-asumsi hipotesis-hipotesis dari model penelitian sebagai berikut :

1. Hipotesis 1 : kemampuan pengurus partai mempunyai pengaruh positif terhadap keberhasilan program partai.
2. Hipotesis 2 : keberhasilan program partai mempunyai pengaruh positif terhadap partai politik.
3. Hipotesis 3 : kemampuan pengurus partai mempunyai pengaruh negatif terhadap partai politik.

hipotesis	Path	Estimasi	Nilai-t	kesimpulan
1	Pengurus → Program	0,76	5,58	Signifikan (Hipotesis 1 diterima)
2	Program → Partai	0,72	2,08	Signifikan (Hipotesis 2 diterima)
3	Pengurus → Partai	-0,23	-0,49	Signifikan (Hipotesis 3 diterima)

16

16. Penskalaan Berdimensi Ganda (Multi dimensional Scaling)

16.1. Pendahuluan

Dalam menyampaikan suatu data atau informasi, seringkali akan lebih mudah dan menarik untuk menampilkannya dalam bentuk gambar. Termasuk dalam menampilkan data-data (atribut) suatu objek. Posisi relatif objek-objek berdasarkan data-data yang dimilikinya, dapat ditampilkan dalam sebuah grafik sehingga lebih mudah dibaca oleh pengguna informasi tersebut. Multidimensional Scalling adalah salah satu tehnik statistika yang dapat diterapkan dalam masalah ini.

Multidimensional Scaling (MDS) merupakan suatu tehnik eksplorasi yang digunakan untuk memvisualisasikan *proximities* (kemiripan/ketakmiripan) dalam ruang dimensi yang rendah. Dari sudut pandang nonteknis, tujuan MDS adalah untuk menyajikan secara visual hubungan beberapa objek dalam sebuah grafik. Interpretasi dari keluaran (*output*) yang dihasilkan MDS dapat mengarah pada pemahaman yang mendasari kedekatan antar objek (entitas). Lebih jauh lagi, dapat dimungkinkan untuk menggabungkan objek-objek yang mirip ke dalam satu kelompok yang sama. MDS merupakan bagian dari analisis multivariat, karena suatu objek

seringkali melibatkan banyak variabel atau peubah yang menjadi atribut-atribut objek tersebut.

16.2. Penskalaan Berdimensi Ganda

16.2.1. Pengertian Penskalaan Berdimensi Ganda

Definisi sederhana dari multidimensional scaling adalah pencarian ruang dimensi yang kecil (pada umumnya menggunakan euklid) yang dapat menyajikan objek-objek sedemikan sehingga jarak antar objek pada ruang dimensi tersebut sesuai dengan jarak asli antar objek yang diamati. Jarak antar objek bukan hanya berarti jarak secara harfiah tetapi dapat pula berarti kemiripan atau ketakmiripan antar objek.

Dari definisi tersebut, kegunaan multidimensional scaling adalah untuk menyajikan objek-objek secara visual berdasarkan kemiripan yang dimiliki. Selain itu kegunaan lain dari teknik ini adalah mengelompokkan objek-objek yang memiliki kemiripan dilihat dari beberapa peubah yang dianggap mampu menggelompokkan objek-objek tersebut.

Ukuran yang digunakan untuk mengukur hubungan antar objek adalah *proximity* yang berarti "kedekatan" objek yang satu dengan objek lainnya. Proximity dapat berupa "kemiripan" $\{s_{rt}\}$ ataupun "ketakmiripan" $\{\delta_{rt}\}$ antar objek, dengan Indeks *r* dan *t* melambangkan objek ke *r* dan objek ke *t* yang dibandingkan. Misalkan suatu himpunan *n* objek yang memiliki ketakmiripan $\{\delta_{rt}\}$ dengan *r,*t = 1, 2, …, *n*. Suatu konfigurasi dari *n* titik dalam ruang dimensi *p* mewakili objek-objek yang diamati, denga jarak antar titik dilambangkan dengan $\{d_{rt}\}$. Masing-masing titik mewakili satu objek dengan titik ke *r* mewakili objek ke *r*. Tujuan dari multidimensional scaling adalah menemukan suatu konfigurasi sedemikian sehingga jarak antar titik sesuai dengan ketakmiripan antar objek.

16.2.2. Jenis-jenis Penskalaan Berdimensi Ganda

Tipe data berdasarkan skala pengukuran dibagi menjadi 4 (empat) tipe, yaitu skala nominal, ordinal, interval dan rasio. Berdasarkan tipe data tersebut, Penskalaan Berdimensi Ganda dibagi menjadi 2 (dua) jenis, yaitu Penskalaan Berdimensi Ganda metrik dan Penskalaan Berdimensi Ganda non-metrik

a. *Penskalaan Berdimensi Ganda Metrik*

Data jarak yang digunakan dalam Penskalaan Berdimensi Ganda metrik adalah data rasio. Penskalaan Berdimensi Ganda metrik digunakan untuk menemukan himpunan titik dalam ruang dimensi *n* dimana masing-masing titik mewakili satu objek sehingga jarak antar titik adalah $d_{rt} \approx f(\delta_{rt})$, dimana *f* adalah fungsi monotonic parametric kontinu. Fungsi ini dapat berupa fungsi identitas maupun fungsi transformasi ketakmiripan menjadi bentuk jarak.

Jenis Penskalaan Berdimensi Ganda metrik yang sering digunakan adalah yang diperkenalkan oleh Young dan Householder pada Tahun 1938. Dalam penskalaan klasik ketakmiripan $\{\delta_{rt}\}$ diperlakukan sebagai jarak euklid. Misalkan koordinat *n* titik dalam ruang euklid dimensi *p* adalah $x_r \ (r = 1, \cdots, n)$, dimana $x_r \left(x_{r1}, \cdots x_{rp}\right)^T$. Jarak euklid antara titik ke *r* dan *t* adalah :

$$d_{rt}^2 = (x_r - x_t)^T (x_r - x_t) \quad \ldots\ldots\ldots\ldots\ldots\ldots (1)$$

Misalkan matriks hasil kali dalam B, dimana

$$[B]_{rt} = b_{rt} = x_r^T x_t$$

Dengan $\sum_{r=1}^{n} x_{ri} = 0 \qquad (i = 1, \cdots, p)$

Untuk mencari B dari persamaan (1) diperoleh :

$$d_{rt}^2 = x_r^T x_r + x_t^T x_t - 2x_r^T x_t \quad \text{........................... (2)}$$

$$\frac{1}{n}\sum_{r=1}^{n} d_{rt}^2 = \frac{1}{n}\sum_{r=1}^{n} x_r^T x_r + x_t^T x_t,$$

$$\frac{1}{n}\sum_{t=1}^{n} d_{rt}^2 = x_r^T x_r + \frac{1}{n}\sum_{t=1}^{n} x_t^T x_t,$$

$$\frac{1}{n^2}\sum_{r=1}^{n}\sum_{t=1}^{n} d_{rt}^2 = \frac{2}{n}\sum_{r=1}^{n} x_r^T x_r \quad \text{.................. (3)}$$

Disubstitusi ke dalam persamaan (2) menjadi

$$b_{rt} = x_r^T x_t$$

$$= -\frac{1}{2}\left(d_{rt}^2 - \frac{1}{n}\sum_{r=1}^{n} d_{rt}^2 - \frac{1}{n}\sum_{s=1}^{n} d_{rt}^2 + \frac{1}{n^2}\sum_{r=1}^{n}\sum_{t=1}^{n} d_{rt}^2\right),$$

$$= a_{rt} - a_{r.} - a_{.t} + a_{..} \quad \text{...(4)}$$

Dimana $a_{rt} = -\frac{1}{2} d_{rt}^2$, dan

$$a_{r.} = n^{-1}\sum_t a_{rt}\ , \qquad a_{.t} = n^{-1}\sum_r a_{rt}\ , \qquad a_{..} = n^{-2}\sum_r\sum_t a_{rt}$$

Matriks A didefinisikan sebagai $[A]_{rt} = a_{rt}$, dan karena hasil kali dalam matriks B adalah

B= HAH ....

(5)

Dimana $\mathrm{H} = \mathrm{I} - \mathrm{n}^{-1}11^{\mathrm{T}}$ dengan 1 = (1, 1, ... , 1)^T adalah vector 1 berukuran *n*.

Matriks hasil kali dalam B dapat juga diekspresikan sebagai $B = \mathrm{XX}^T$, dimana $\mathrm{X} = [\mathrm{x}_1, \cdots, \mathrm{x}_\mathrm{n}]^\mathrm{T}$ adalah matriks koordinat berukuran *n x p*. Pangkat dari matriks B, r(B) adalah

$$r(\mathrm{B}) = r(\mathrm{XX}^\mathrm{T}) = r(\mathrm{X}) = p.$$

Sekarang B adalah matriks yang simetrik, semi definit positif dan berpangkat *p*, sehingga memiliki *p* akar ciri nonnegative dan *n-p* akar ciri 0.

Matriks B kemudian ditulis dalam bentuk dekomposisi spectral, $B = V\Lambda V^T$, dimana $\Lambda=\mathrm{diag}(\lambda_1,\lambda_2, \ldots , \lambda_n)$, yaitu matriks diagonal dari akar ciri $\{\lambda_i\}$ matriks B, dan $V = [v_1, \ldots , v_n]$, yaitu matriks vektor akar ciri yang dinormalkan menjadi $v_i^T v_i = 1$. Akar ciri yang diperoleh kemudian disusun menjadi $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n \geq 0$.

Karena memiliki *n – p* akar ciri 0, maka matriks B dapat ditulis kembali sebagai $B = V_1\Lambda_1V_1^T$, dimana $\Lambda_1 = \mathrm{diag}(\lambda_1, \ldots, \lambda_p)$, $V_1 = [v_1, \ldots , v_p]$.

Karena $B = XX^T$, maka koordinat matriks X adalah $X = V_1\Lambda_1^{1/2}$, dimana $\Lambda_1^{1/2} = \mathrm{diag}(\lambda_1^{1/2}, \ldots,\lambda_p^{1/2})$.

Penggunaan classical scaling biasanya lebih banyak mengenai ketakmiripan $\{\delta_{rt}\}$ dibandingkan dengan jarak euklid sebenarnya antar titik $\{d_{rt}\}$.

Menurut Mardia *et al*, jika B adalah semi definit positif berpangkat *p*, maka $B = V\Lambda V^T = XX^T$, dimana $X = [x_r]^T$, $x_r = \lambda^{1/2}v_r$. Jarak antara titik ke *r* dan ke *t* dari konfigurasi adalah $(x_r - x_s)^T(x_r - x_t)$, diperoleh

$$\begin{aligned}(x_r - x_t)^T(x_r - x_t) &= x_r^T x_r + x_t^T x_t - 2x_r^T x_t \\ &= b_{rr} + b_{tt} - 2b_{rt} \\ &= a_{rr} + a_{tt} - 2a_{rt} = -2a_{rs} = \delta_{rt}^2\end{aligned}$$

Dengan δ_{rt}^2 sama dengan jarak antara titik *r* ke *t* dalam ruang euklid.

Jika koefisien ketakmiripan menyebabkan matriks B tidak semi definit positif, suatu konstanta dapat ditambahkan pada semua koefisien ketakmiripan (kecuali $\{\delta_{rr}\}$) sehingga matriks B menjadi matriks semi definit positif. Bentuk koefisien ketakmiripan yang baru menjadi $\delta_{rt}^{'} = \delta_{rt} + c(1 - \delta^{rt})$, dimana c adalah suatu konstanta dan δ^{rt} adalah kronecker

delta ($\delta^{rt} = 1$ jika *r* = *t* dan 0 untuk lainnya, tidak ada hubungan dengan δ_{rt}).

Permasalahan berikutnya adalah menentukan jumlah dimensi yang diperlukan untuk menampilkan koefisien ketakmiripan $\{\delta_{rt}\}$. Jika B adalah matriks semi definit positif maka jumlah akar ciri yang tak nol menujukkan jumlah dimensi yang diperlukan. Jika B bukan matriks semi definit positif maka jumlah akar cirri yang positif menunjukkan jumlah dimensi yang tepat. Jumlah dimensi tersebut merupakan jumlah dimensi maksimal yang diperlukan, sedangkan untuk lebih praktisnya lebih baik memilih dimensi yang lebih kecil.

Dari persamaan (3), jumlah kuadrat jarak antar titik dalam ruang adalah

$$\frac{1}{2}\sum_{r=1}^{n}\sum_{s=1}^{n} d_{rt}^2 = n\sum_{r=1}^{n} x_r^T x_r = n \text{ tr B} = n\sum_{r=1}^{n-1} \lambda_i$$

Suatu ukuran proporsi variasi yang dijelaskan dengan menggunakan hanya dimensi *p* adalah

$$\frac{\sum_{i=1}^{p} \lambda_i}{\sum_{i=1}^{n-1} \lambda_i}$$

Jika B bukan matriks semi definit positif, ukuran tersebut dimodifikasi menjadi

$$\frac{\sum_{i=1}^{p} \lambda_i}{\sum(akar\ ciri\ positif)}$$

Ukuran tersebut dapat digunakan untuk memilih jumlah dimensi (*p*) yang akan digunakan.

Secara singkat algoritma classical scaling sebagai berikut :

1. Menentukan koefisien ketakmiripan $\{\delta_{rt}\}$
2. Mencari matriks A = $\left[-\frac{1}{2}\delta_{rt}^{2}\right]$
3. Mencari matriks B = $[a_{rt} - a_{r.} - a_{.s} + a_{..}]$
4. Mencari akar cirri λ_1 , … , λ_{n-1} dan vektor ciri v_1, … , v_{n-1} yang kemudian dinormalkan sehingga $v_i^T v_i = \lambda_i$. Jika B tidak semi definit positif (beberapa akar ciri bernilai negatif), maka terdapat 2 pilihan, pilihan 1 adalah membuang akar ciri yang bernilai negatif dan melanjutkan proses. Pilihan ke 2 adalah menambahkan suatu konstanta *c* pada koefisien ketakmiripan sebagai berikut $\delta_{rt}^{'} = \delta_{rt} + c(1 - \delta^{rt})$ dan kembali ke langkah 2.
5. Memilih jumlah dimensi yang tepat. Dapat menggunakan $\sum_i^p \lambda_i / \sum(akar\ ciri\ positif)$.
6. Menentukan koordinat n titik pada ruang euklid dimensi *p* dengan $x_{ri} = v_{ir}$ $(r = 1, \ldots, n\,; i = 1, \ldots, p$.

b. *Penskalaan Berdimensi Ganda Non Metrik*

Data jarak yang digunakan dalam Penskalaan Berdimensi Ganda non metrik adalah data yang dianggap bertipe ordinal. Untuk Penskalaan Berdimensi Ganda non metrik, fungsi transformasi hanya mempunyai batasan $\delta_{rt} < \delta_{r't'} \Rightarrow f(\delta_{rt}) \leq f(\delta_{r't'})$ untuk semua $1 \leq r, t, r^{'}, t^{'} \leq n$. Suatu fungsi stress sebagai berikut :

$$Stress = \sqrt{\frac{\sum_{r,t}\left(d_{rt} - \hat{d}_{rt}\right)^2}{\sum_{r,t} d_{rt}^2}}$$

Fungsi ini mengambil 1 = *r* < *s* = *n*, karena $\delta_{st} = \delta_{ts}$ untuk semua *r, t*. Meminimumkan Fungsi stress dengan memperhatikan $\{d_{rt}\}$ dan $\{\hat{d}_{rt}\}$ menggunakan regresi isotonic.

Dalam berbagai literatur regresi isotonic mempunyai arti regresi kuadrat terkecil monoton utama dari $\{d_{rt}\}$ pada $\{\delta_{rt}\}$.

Ide untuk menemukan tampilan dari objek-objek dalam titik pada ruang dimensi *q* sedemikian sehingga nilai stress sekecil mungkin. Kruskal menyarankan bahwa dengan nilai stress dapat kebaikan suai dari hubungan monotonic antara kemiripan dan jarak akhir, dengan ketentuan nilai sebagai berikut

STRESS (%)	Kesuaian
20	Buruk
10	Cukup
5	Bagus
2.5	Sangat bagus
0	Sempurna

Langkah pertama meminimumkan Stress adalah menempatkan semua koordinat titik dalam X dalam suatu vector $x = (x_{11}, \dots, x_{1p}, \dots, x_{np})^T$ sehingga stress sebagai fungsi dari x yang diminimumkan dengan cara iteratif. Metode penurunan tajam digunakan, sehingga jika x_m adalah vector koordinat setelah iterasi ke *m*

$$x_{m+1} = x_m - \frac{\frac{\partial S}{\partial x} x \; sl}{\left|\frac{\partial S}{\partial x}\right|}$$

$$\frac{\partial S}{\partial x_{ui}} = S \sum_{r,t} (\partial^{ru} - \partial^{tu}) \left[\frac{d_{rt} - \hat{d}_{rt}}{S^*} - \frac{d_{rt}}{T^*}\right] x \; \frac{|x_{ri} - x_{ti}|^{\lambda - 1}}{d_{rt}^{\lambda - 1}}$$

signum $(x_{ri} - x_{ti})$

Berikut adalah teknik iteratif Kruskal yang digunakan untuk menemukan konfigurasi dengan nilai stress minimum :

1. Memilih suatu konfigurasi awal.
2. Menormalisasi konfigurasi untuk mendapatkan centroid pada data aslinya dan jarak kuadrat tengah unit dari data aslinya δ_{rt}. Hal ini dilakukan karena stress invariant terhadap translasi, dilatasi seragam.
3. Tentukan δ_{rt} dari konfigurasi yang telah dinormalkan.
4. Menyesuaikan $\hat{d}_{rt}$. Regresi kuadrat terkecil monotonic dari d_{rt} pada δ_{rt} membagi δ_{rt} menjadi blok-blok dimana $\hat{d}_{rt}$ konstan, dan sama dengan nilai tengah dari d_{rt}. Untuk menemukan partisi δ_{rt} yang tepat, partisi terbaik digunakan adalah yang memiliki N blok dengan masing-masing berisi suatu $\delta_i$menggunakan notasi alternatif. Jika partisi ini memiliki $d_1 \leq d_2 \leq \cdots \leq d_N$, dan $\hat{d}_i = d_i$ maka partisi ini merupakan partisi yang tepat. Jika tidak demikian maka blok yang berurutan digabung dimana $\delta_i > \delta_{i+1}$ dan $\hat{d}_i = \hat{d}_{i+1} = (d_i + d_{i+1})/2$. Blok terus-menerus digabung dan $\hat{d}_i$ baru selalu diperoleh hingga partisi yang dibutuhkan tercapai.
5. Temukan gradient $\frac{\partial S}{\partial x}$. Jika $\left|\frac{\partial S}{\partial x}\right| < \epsilon$, dimana ϵ adalah nilai yang sangat kecil. Jika suatu konfigurasi dengan stress minimum diperoleh maka proses iteratif berhenti.
6. Temukan panjang sl.

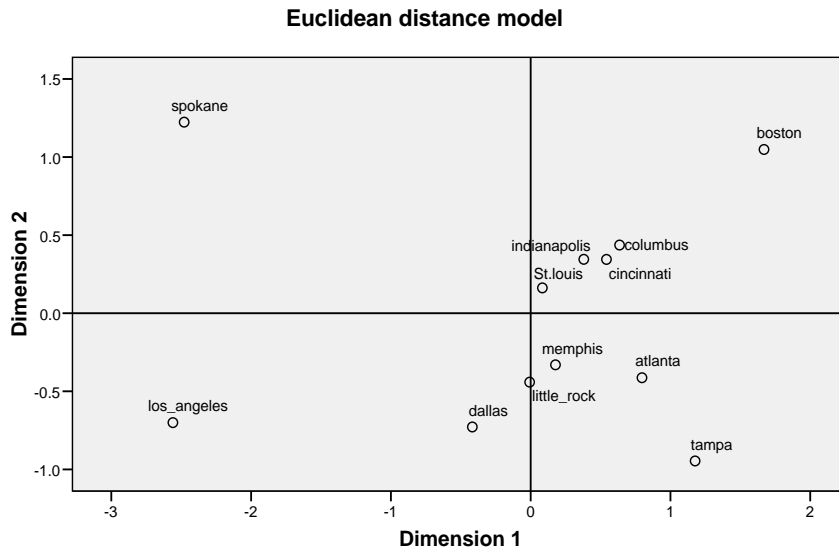
7. Temukan konfigurasi yang baru, yaitu $x_{n+1} = x_n - sl\frac{\frac{\partial s}{\partial x}}{\left|\frac{\partial s}{\partial x}\right|}$

8. Kembali ke langkah 2.

16.3. Ilustrasi

Berikut adalah data jarak antar pasangan kota terpilih di US.

	Atlanta	Boston	Cincinnati	Coloumbus	Dallas	Indianapolis	Little Rock	Los Angeles	Memphis	St. Louis	Spokanes	Tampa
Atlanta	0	0	0	0	0	0	0	0	0	0	0	0
Boston	1068	0	0	0	0	0	0	0	0	0	0	0
Cincinnati	461	867	0	0	0	0	0	0	0	0	0	0
Coloumbus	549	769	107	0	0	0	0	0	0	0	0	0
Dallas	805	1819	943	1050	0	0	0	0	0	0	0	0
Indianapolis	508	941	108	172	882	0	0	0	0	0	0	0
Little Rock	505	1494	618	725	325	562	0	0	0	0	0	0
Los Angeles	2197	3052	2186	2245	1403	2080	1701	0	0	0	0	0
Memphis	366	1355	502	586	464	436	137	1831	0	0	0	0
St. Louis	558	1178	338	409	645	234	353	1848	294	0	0	0
Spokanes	2467	2747	2067	2131	1891	1959	1988	1227	2042	1820	0	0
Tampa	467	1379	928	985	1077	975	912	2480	779	1016	2821	0

Derived Stimulus Configuration

Tunjukkan posisi kota-kota di atas menggunakan penskalaan berdimensi ganda.

Dengan menggunakan SPSS 15, menghasilkan output :

Dari gambar di atas tampak bahwa kota Indianapolis, Columbus, Cincinnati, St.Louis saling berdekatan dibandingkan dengan kota-kota lain dan tampak berada dalam wilayah yang sama. Jarak paling jauh adalah antara kota Boston dan Los Angeles serta antara Spokane dan Tampa.

Stress and Fit Measures

Normalized Raw Stress	.00047
Stress-I	.02168(a)
Stress-II	.03934(a)
S-Stress	.00117(b)
Dispersion Accounted For (D.A.F.)	.99953
Tucker's Coefficient of Congruence	.99976

PROXSCAL minimizes Normalized Raw Stress.

a Optimal scaling factor = 1.000.

b Optimal scaling factor = 1.001.

Dari nilai Koefisien Tucker yang dihasilkan sebesar 0.99976 hampir 100% yang berarti model multidimensional scaling sempurna. Hal ini dapat terjadi karena data yang digunakan adalah data jarak antar kota yang biasa digunakan dalam 2 dimensi. Ini didukung pula dengan nilai Normalized Raw Stress yang hanya sebesar 0.00047, yang berarti hampir 100% data dapat dijelaskan oleh multidimensional scaling.

Decomposition of Normalized Raw Stress

		Source	Mean
		SRC_1	SRC_1
Object	atlanta	.0004	.0004
	boston	.0006	.0006
	cincinnati	.0001	.0001
	columbus	.0001	.0001
	dallas	.0002	.0002
	indianapolis	.0002	.0002
	little_rock	.0003	.0003
	los_angeles	.0013	.0013
	memphis	.0001	.0001
	St.louis	.0001	.0001
	spokane	.0019	.0019
	tampa	.0004	.0004
Mean		.0005	.0005

Tabel ini menjelaskan tentang objek-objek yang memberikan kontribusi paling besar terhadap stress. Berdasarkan tabel tersebut kota Spokane memberikan kontribusi terbesar terhadap nilai stress, yaitu sebesar 0.0019.

Iteration History

Iteration	Normalized Raw Stress	Improvement
0	.23972(a)	
1	.02270	.21701
2	.01611	.00660
3	.01199	.00411
4	.00900	.00299
5	.00678	.00222
6	.00512	.00166
7	.00387	.00125
8	.00293	.00094
9	.00223	.00070
10	.00171	.00052
11	.00133	.00039
12	.00104	.00029
13	.00083	.00021
14	.00067	.00016
15	.00056	.00012
16	.00047	.00009(b)

a Stress of initial configuration: simplex start.

b The iteration process has stopped because Improvement has become less than the convergence criterion.

Tabel iterasi history menunjukkan proses iterasi hingga diperoleh nilai stress yang konvergen. Untuk analisis data digunakan criteria iterasi untuk kekonvergenan stress sebesar 0.0001.

16.4. Aplikasi SAS

Untuk melakukan analisis MDS, kita menggunakan nilai-nilai yang menggambarkan tingkat kemiripan atau pun tingkat ketakmiripan antar objek. Nilai-nilai tersebut sering disebut sebagai *proximity*

- *similarity* (kemiripan) jika semakin besar nilai menunjukkan bahwa objek nya lebih mirip
- *dissimilarity* (ketakmiripan) jika semakin besar nilainya menunjukkan bahwa objeknya semakin tidak mirip

Jika menggunakan SAS, prosedur yang digunakan untuk penskalaan dimensi ganda adalah PROC MDS.

Sebagai ilustrasi penggunaan PROC MDS pada SAS, misalnya kita memiliki data tentang jarak antar kota, kita bisa membuat peta lokasi berdasarkan data tersebut dengan menggunakan multidimensional scaling. Berikut ini akan kita lakukan analisis terhadap data jarak beberapa kota di Amerika.

Penyiapan datanya adalah sebagai berikut :

```
data   city;
title 'Analysis of Flying Mileages Between Ten U.S. Cities';
input  (atlanta chicago denver houston losangeles miami newyork
sanfran seattle washdc) (5.) @56 city $15.;
datalines;
    0                                                Atlanta
  587    0                                           Chicago
 1212  920    0                                      Denver
  701  940  879    0                                 Houston
 1936 1745  831 1374    0                            Los Angeles
  604 1188 1726  968 2339    0                       Miami
  748  713 1631 1420 2451 1092    0                  New York
 2139 1858  949 1645  347 2594 2571    0             San Francisco
 2182 1737 1021 1891  959 2734 2408  678    0        Seattle
  543  597 1494 1220 2300  923  205 2442 2329    0    Washington D.C.
;
```

Proximity yang ada adalah jarak antara dua kota. Karena jarak bersifat simetrik, jarak antara Chicago ke Atlanta sama dengan jarak Atlanta ke Chicago, maka hanya setengah dari tabel saja yang perlu dimasukkan.

Selanjutnya kita gunakan PROC MDS dari SAS/STAT untuk membuat data set yang berisi koordinat titik-titik kota yang akan di plot.

```
proc mds data=city level=absolute out=out;
   id city;
run;
```

Beberapa option pernyataan pada PROC MDS yang digunakan adalah

LEVEL= menyebutkan sakal pengukuran data. Karena kita menyebutkan LEVEL=ABSOLUTE jarak antara koordinat yang di plot akan sama dengan jarak aslinya.

OUT= membuat data set yang berisi koordinat dari kota-kota dan statistik lain hasil analisis

ID = merupakan pernyataan yang menyatakan peubah mana yang dijadikan nama objek.

Option OUT= akan menghasilkan gugus data hasil analisis menggunakan PROC MDS yang memuat peubah DIM1 dan DIM2, yang nilainya menunjukkan koordinat titik yang ingin diplot. Jika kita mebuat plot dari koordinat itu :

- tambahkan label pada plot sebagai identitas titik
- gambar sumbu vertikal dan horizontal dengan skala yang sama. Jika kita lakukan dengan cara ini, maka jarak terlihat dengan benar.

Untuk membuat plot hasil keluaran PROC MDS ini bisa digunakan alat-alat yang ada di SAS System, antara lain :

- makro %PLOTIT, makro yang disediakan oleh SAS/STAT sample library
- PROC PLOT, prosedur yang disediakan oleh software SAS

Menampilkan plot dengan makro %PLOTIT

Kita bisa menggunakan makro %PLOTIT untuk membuat plot dari hasil keluaran PROC MDS. Makro ini merupakan bagian dari SAS/STAT sample library dan bisa ditemukan pada SAS Release 6.08 atau yang lebih baru.

Makro %PLOTIT bekerja dengan menampilkan titik-titik dan label sehingga mereka tidak saling tumpang tidih di plot yang dihasilkan.

Pada contoh di atas, perintah yang diberikan adalah

```
%plotit(data=out, datatype=mds, labelvar=city, plotvars
= dim2 dim1, labfont=swissb);

run;
```

Nilai nilai masukan untuk makro ini adalah :

DATA=	menyebutkan data yang menjadi input makro, yaitu data yang diperoleh dari option OUT= pada PROC MDS
PLOTVARS=	menyebutkan peubah yang dijadikan sumbu vertical dan horizontal pada plot
DATA TYPE=	menyebutkan tipe analisis data yang digunakan untuk menghasilkan data tersebut. DATATYPE=MDS berarti menyatakan bahwa data dihasilkan dari multidimensional scaling.
LABELVAR=	menyebutkan peubah apa yang dijadikan label titik di plot. Pada contoh ini digunakan peubah CITY.

Plot tersebut menunjukkan posisi relatif masing-masing kota. Karena kita menspesifikasi LEVEL=ABSOLUTE dalam tahapan di PROC MDS, skala pada salb sumbu sama dengan skala pada data yang diinput.

Karena orientasi arah pada plot adalah semabrang, maka arah utara bukan pada bagian atas grafik. Pada plot ini, arah utara ada di bagian dasar grafik. PROC MDS menempatkan titik asal di tengah-tengah konfigurasi.

Walaupun tidak harus dan tidak mengubah interpretasi, kadangkala diperlukan mengganti arah grafik. Misalnya pada grafik di atas kita bisa mengalikan peubah DIM2 dengan –1 dan selanjutnya membuat plot dengan DIM1.

```
data out;
  set out;
  dim2 = -1* dim2;
run;
%plotit(data=out,   datatype=mds,   labelvar=city,   plotvars   =   dim2   dim1,
labfont=swissb);
run;
```

sekarang arah utara ada di bagian atas grafik.

17

17.Analisis Konjoin (Conjoint Analysis)

17.1. Pendahuluan

Analisis konjoin (*conjoint analysis*) merupakan suatu metode analisis dalam analisis multivariate, analisis ini mulai diperkenalkan pada tahun 1970-an (Cattink and Wittink, 1992). Analisis ini biasa diterapkan pada market riset dan studi pengembangan produk. Analisis konjoin adalah sebuah teknik guna mengukur preferensi konsumen terhadap produk atau jasa. Analisis konjoin berdasarkan pada subjektifitas konsumen terhadap beberapa kombinasi fitur yang ditawarkan. Subjektifitas konsumen ini diukur melalui peringkat (rank) atau skore (skala likert). Hasil analisis konjoin berupa informasi kuantitatif yang dapat memodelkan preferensi konsumen untuk beberapa kombinasi fitur produk. Sebagai contoh, ketika konsumen hendak membeli komputer mungkin akan memeriksa himpunan atribut untuk memilih produk mana yang paling cocok dengan kebutuhannya. Konsumen mungkin akan mempertimbangkan unsur kecepatan, merek *motherboard*, daya tampung memori, jenis *VGA card* dan lain-lain. Pengukuran dan analisis dalam penelitian pemasaran untuk memilih suatu produk biasanya dilakukan dengan menggunakan analisis conjoin.

Dalam prosesnya analisis konjoin akan memberikan ukuran kuantitatif terhadap tingkat kegunaan (*utility*) dan kepentingan relatif (*relatif importance*) suatu atribut dibandingkan dengan atribut lain. Hal ini dilakukan melalui pertimbangan psikologis atau preferensi konsumen

(Green & Tull, 1988). Lebih lanjut, nilai-nilai ini dapat digunakan untuk membantu menyeleksi atribut-atribut suatu produk yang akan ditawarkan.

Tujuan analisis konjoin adalah memperoleh skor kegunaan (*utility*) yang dapat mewakili kepentingan setiap aspek produk, sehingga dari skor tersebut dapat ditarik kesimpulan tentang atribut apa yang paling dipertimbangkan konsumen dalam memilih produk. Dalam pemasaran teknik analisis konjoin biasanya digunakan untuk hal-hal berikut :

1. Menentukan tingkat kepentingan relative atribut-atribut pada proses pemilihan yang dilakukan oleh konsumen .*Output* baku dari analisis conjoin terdiri dari kepentingan relative dari timbangan yang diturunkan untuk semua atribut yang dipergunakan untuk membangun stimulus yang diperuntukan dalam tugas evaluasi. Kepentingan relative timbangan (*weights*) menunjukan atribut mana yang penting dalam mempengaruhi pilihan pelanggan.
2. Membuat estimasi pangsa pasar suatu produk tertentu yang berbeda tingkat atributnya. The utilities yang diturunkan dari analisis conjoin bisa dipergunakan sebagai input ke dalam suatu pilihan simulator untuk menentukan sumbangan pilihan, kemudian pangsa pasar dengan berbagai jenis merk.
3. Menentukan komposisi produk yang paling disukai konsumen. Feature dari merk dapat dibuat bervariasi dinyatakan dalam tingkatan/level atribut dan utilities yang bersangkutan ditentukan. Feature dari merk yang menghasilkan utility tertinggi menunjukan komposisi merk yang paling disenangi.
4. Membuat segmentasi pasar yang didasarkan pada kemiripan preferensi terhadap tingkat-tingkat atribut. Fungsi parth-worth diturunkan untuk atribut, mungkin dipergunakan sebagai basis (dasar) untuk mengelompokkan (clustering) responden untuk mencapai segmen dreferensi yang homogeny.

Sementara manfaat yang dapat diambil dari penggunaan analisis konjoin adalah produsen dapat mencari solusi kompromi yang optimal dalam merancang atau mengembangkan suatu produk. Menurut Green dan Krieger (1991) analisis ini dapat juga dimanfaatkan untuk :

1. Merancang harga
2. Memprediksi tingkat penjualan atau penggunaan produk (market share), uji coba konsep produk baru.
3. Segmentasi preferensi
4. Merancang strategi promosi

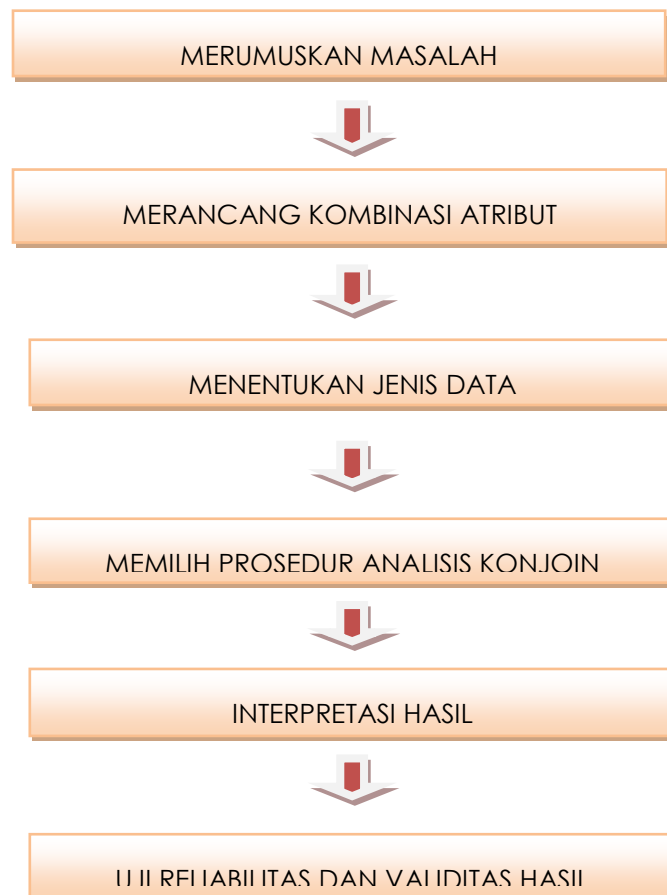
17.2. Statistik Dalam Analisis Konjoin

Statistik yang sering muncul dalam analisis conjoin adalah :

- *Part-worth function* (*utility function*) ialah kegunaan (*utility*) yang dikaitkan oleh pelanggan pada tingkatan/*level* setiap atribut.
- *Relatif importance weight* ialah nilai yang bisa menunjukan atribut mana yang penting di dalam pilihan pelanggan.
- *Atribut level* ialah nilai yang menunjukan tingkatan setiap atribut.
- *Full profiles* atau *complete profiles* dan merk yang dibentuk dinyatakan dalam semua atribut dengan menggunakan atribut level yang ditentukan oleh desain.
- *Pairwise table*, responden mengevaluasi dua atribut pada saat yang sama, sampai semua pasangan atribut sudah dievaluasi.
- *Cyclical design* ialah desain yang dipergunakan untuk mengurangi banyaknya pasangan yang harus diperbandingkan.
- *Factional factorial design* ialah desain yang dipergunakan untuk mengurangi banyaknya profil stimulus yang dievaluasi di dalam pendekatan profil penuh.
- *Orthogonal arrays* ialah sebuah kelas desain factorial yang memungkinkan untuk membuat perkiraan yang efesien dari seluruh pengaruh utama.

Internal validity meliputi korelasi antara evaluasi untuk *hold out* yang diprediksi atau validasi stimulus dengan hasil yang diperoleh dari para responden.

1. **Tahapan Analisis Konjoin**

   Tahapan yang umumnya dilakukan dalam merancang dan melaksanakan analisis konjoin secara umum ditampilkan dalam Gambar berikut:

Gambar 1. Tahapan Analisis Konjoin

1. Perumusan masalah

   Di dalam merumuskan analisis conjoin, peneliti harus mengenali/mengidentifikasi atribut dengan tingkatan/level masing-masing dipergunakan untuk membentuk stimulus. Level atribut menunjukkan nilai yang diasumsikan oleh atribut. Atribut yang dipilih harus dapat mempengaruhi preferensi dan pilihan pelanggan. Atribut bisa diidentifikasi melalui diskusi dengan manajemen dan tenaga ahli, menganalisis data sekunder, riset kualitatif dan lain-lain. Banyaknya tingkatan atribut menentukan banyaknya parameter yang akan diperkirakan dan juga mempengaruhi banyaknya stimulus yang akan dievaluasi oleh responden. Untuk meminimumkan tugas evaluasi responden, peneliti harus bisa memperkirakan parameter seakurat mungkin, perlu membatasi banyaknya tingkatan/level dari atribut. Peneliti harus memperhitungkan level atribut yang lazim atau umum berlaku di masyarakat . Menggunakan level atribut di luar kisaran yang tercemin di dalam pasar akan mengurangi believability dari tugas evaluasi, akan tetapi akan meningkatkan akurasi dengan mana parameter akan diestimasi.

2. Merancang kombinasi atribut (stimuli)

   Terdapat beberapa ketentuan dalam memilih metode yang akan digunakan dalam merancang kombinasi taraf dari atribut (Hair et al.. 1998), yaitu :

   a. Choice-Based conjoin (CBC). Digunakan apabila jumlah atribut ≤ 6

      Metode ini mulai popular pada awal tahun 1990-an, dan saat ini banyak digunakan serta mendapat perhatian yang sangat besar oleh kalangan peneliti dan praktisi pemasaran. Menurut Hair et al. (1998) keunggulan utama CBC dibandingkan metode

lain adalah prosedur pengumpulan datanya secara langsung mencerminkan prilaku pasar. Metode CBC tidak tepat digunakan dalam penelitian dengan jumlah atribut yang banyak. Green dan Srinivasan (1990), menyatakan bahwa 6-10 atribut adalah jumlah maksimum atribut yang dapat menggunakan konsep full-profile dalam analisis conjoin.

b. Traditional Conjoin, digunakan apabila jumlah atribut < 10
c. Adaptive Conjoin Analysis (ACA), digunakan apabila jumlah atribut ≥ 10.

Adapun pendekatan yang sering digunakan dalam merancang kombinasi atribut yang sering digunakan adalah :

a. Kombinasi berpasangan (*pairwise combination*)
   Responden diminta untuk mengevaluasi pasangan-pasangan atribut secara bersamaan.
b. Kombinasi lengkap (*full profile*)
   Responden diminta mengevaluasi semua kombinasi stimuli yang muncul.

Jika jumlah kombinasi terlalu banyak maka dilakukan pengurangan jumlah kombinasi atribut (stimuli) tersebut. Salah satu cara dengan orthogonal array sehingga akan diperoleh suatu kombinasi atribut yang hanya mengukur efek utamanya saya , sementara interaksi antar atribut tidak terukur atau diabaikan dan jumlah stimuli yang terbentuk akan berkurang.

3. Menentukan jenis data yang diperlukan

Jenis-jenis data pada analisis konjoin, yaitu :

- Nonmetrik (data dalam bentuk nominal atau ordinal atau ordinari)
  Membuat ranking atau mengurutkan stimuli yang telah dibuat pada tahap sebelumnya. Pengurutan ini biasanya

dimulai dari stimuli yang paling disukai sampai pada stimuli yang paling tidak disukai.

- Metrik (data berskala interval atau rasio)

  Memberikan rating atau nilai terhadap masing-masing stimuli.

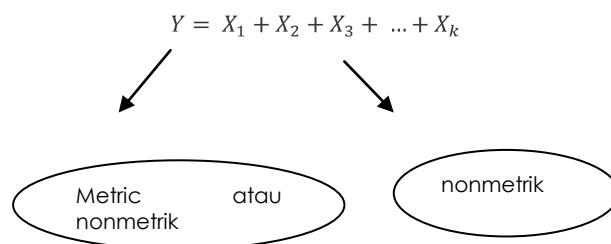
Dalam konjoin analisis yang berperan sebagai vaiabel tak bebas (*dependent variable*) umumnya adalah preferensi atau keinginan untuk membeli. Oleh karenanya pemberikan urutan maupun penilaian didasarkan atas kedua hal tersebut. Namun demikian karena proses konjoin bersifat sangat fleksibel maka pembelian secara aktual maupun pilihan dapat berlaku sebagai variabel tak bebas.

Pemberian nilai atau *rating* dapat dilakukan melalui beberapa cara:

a. Menggunakan skala *lickert* mulai dari 1 hingga 9 (1 = Paling tidak disukai, 9 = Paling disukai).
b. Menggunakan nilai rangking terbalik, artinya untuk stimuli yang paling disukai diberi nilai tertinggi setara dengan jumlah stimulinya, sedangkan stimuli yang paling tidak disukai diberi nilai 1.

4. Memilih prosedur analisis conjoin

Konjoin termasuk dalam *Multivariate Dependence Method*, yaitu hubungan antar variablenya depedensi. Hubungan depedensi yaitu jika variable-variabel yang diteliti secara teoritis dapat dipisahkan kedalam variable-variabel respon dan variable penjelas (Santoso, 2002) dengan model sebagai berikut :

$$Y = X_1 + X_2 + X_3 + \dots + X_k$$

Variabel independen (X) adalah factor dan berupa data nonmetrik. Termasuk disini adalah bagian dari factor (level), sedangkan variable dependen (Y) adalah pendapat keseluruhan (overall preference) dari seorang responden terhadap sekian factor dan level pada sebuah produk.

Secara umum model dasar analisis Conjoin dapat ditulskan dalam bentuk sebagai berikut:

$$U(X) = \sum_{i=1}^{m} \sum_{j=1}^{k_i} \beta_{ij} x_{ij}$$

Keterangan :

- *U(X)* = Utility total
- *βij* = *Part worth* atau nilai kegunaan dari atribut ke-*i* taraf ke-*j*.
- *kI* = Taraf ke-*j* dari atribut ke-i
- *m* = Jumlah atribut
- *xij* = *Dummy variable* atribut ke-i taraf ke-*j*. (bernilai 1 bila taraf yang berkaitan muncul dan 0 bila tidak)

Fungsi kegunaan (utility function) yang nantinya akan diduga sangat dipengaruhi oleh model yang sangat dipengaruhi oleh model yang akan digunakan di bawah ini :

1. Factor diskret

$$\widehat{u_{jk}} = \begin{cases} \hat{\alpha}_{jk}, & untuk\ k = 1, 2, \dots, m_j - 1 \\ -\sum_{j=1}^{m_{j-1}} \hat{\alpha}_{jk}, & untuk\ k = m_j \end{cases}$$

2. Factor linier $\hat{u}_{jk} = \hat{\beta}_j x_k$
3. Factor ideal-point $\hat{u}_{jk} = {}_{\varepsilon}\hat{\gamma}_{ji} z_{jk} + \hat{\gamma}_{j2} z^2_{jk}$

Salah satu metode yang dapat digunakan untuk menyelesaikan model dari analisis conjoin adalah metode regresi dengan variable dummy.

Untuk atribut ke-j dengan k_j level, variable dummynya adalah :

Tabel.1. Variabel Dummy Atribut ke-j dan level k_j

Level	X_1	X_2	…	x_{kj-1}
1	1	0	. . .	0
2	0	1	. . .	0
3	0	0	. . .	0
. . .	. . .	. . .	. . .	. . .
k_j - 1	0	0	. . .	1
k_j	0	0	. . .	0

Langkah yang paling penting dalam analisis conjoin adalah mengestimasi kegunaan (utility function) atau tingkat kepentingan relative individu (individual level part worth)

Untuk mendapatkan nilai-nilai $u_{jk_{ji}}$ tersebut, langkah yang harus dilakukan adalah mengestimasi model dasar analisis conjoin dengan persamaan regresi linier ganda dengan variable dummy. Maka persamaan regresinya adalah :

$$Y_i = \beta_0 + \beta_{1i} + \beta_{2i} + \ldots + \beta_k X_{ki} + \varepsilon_i$$

Untuk menaksir parameter pada persamaan di atas maka akan digunakan metode kuadrat terkecil.

Pentingnya suatu atribut, misalkan $RANGE_i$ dinyatakan dalam kisaran Part Worth melintasi level dari atribut, yaitu :

$RANGE_i = \left\{\max\left(u_{jk_{ji}}\right) - \min\left(u_{jk_{ji}}\right)\right\}$, untuk setiap i

Pentingnya suatu atribut digunakan untuk meyakinkan kepentingan relative disimbolkan dengan IMP yang ditentukan melalui formula berikut :

$$IMP_i = \frac{RANGE}{\sum_{I=1}^{P} RANGE_i} \; x\; 100\%$$

Setelah didapatkan nilai-nilai $u_{jk_{ji}}$, maka kisaran part worth $RANGE_{id}$ an timbangan kepentingan relative IMP_i akan diperoleh. Kisaran part worth dan timbangan kepentingan relative ini memberikan dasar untuk menginterpretasikan hasil. Angka IMP_i yang terbesar menunjukan preferensi terbesar terhadap level-level pada atribut tertentu.

17.3. Ilustrasi

17.3.1. Metric Conjoint Analysis

Contoh ini memberikan gambaran sederhana analisis konjoin / *conjoint analysis*(CA) dengan subjek tunggal.Subjek ditanya untuk menilai 8 jenis permen coklat.adapun atribut permen coklat yaitu: dari segi bahan dasar,ada jenis coklat hitam atau coklat susu (*dark/milk*), dari segi isi tengah permen ada permen kenyal atau lembut (*chewy/soft*), dan terakhir dari segi ada tidaknya kacang dalam permen(*nuts/no nuts*). Skala penilaian dari 1-9, dengan 1 adalah pilihan yang paling tidak disukai hingga 9 adalah pillihan yang paling disukai.CA digunakan untuk menentukan atribut utama dan bagian penting dari suatu permen coklat.

Setelah pengumpulan, atribut dan data penilaian dimasukkan kedalam SAS data set.

```
title'Preferensi permen coklat';

data cok;

          input coklat $ isi $ kacang $& rating;

          datalines;
dark chewy nuts     7
dark chewy no nuts  6
dark soft  nuts     6
dark soft  no nuts  4
milk chewy nuts     9
milk chewy no nuts  8
milk soft  nuts     9
milk soft  no nuts  7
;
odsexclude notes mvanova anova;
proctransregutilitiesseparators=',' short;
          title2'Metric Conjoint Analysis';
          modelidentity(rating)=class(coklat isi kacang / zero=sum);
          run;
```

dalam hal ini PROC TRANSREG digunakan dalam CA.output dari metrik CA dimunculkan oleh pernyataan utilitiespada pernyataan PROC .pernyataanseparators=', ' mengakibatkan label terdiri dari nama variabel, koma, spasi lalu atribut dari variabel, kita menggunakan short option untuk menekan itrasi. Pernyataanodsexclude notes mvanova anova; adalah untukmengecualikan informasi ANOVA untuk menghemat output. Analisis variabel, transformasi dari masing masing variabel, transformasi dari pilihan tertantu dinyatakan dalam pernyataan model.

Pernyataan identity(rating) meminta transformasi identitas dari variabel bebas respon “rating”.

Output

```
                              The TRANSREG Procedure

                               Dependent Variable Identity(rating)

                                     Class Level Information

                                Class     Levels    Values

coklat         2    dark  milk

isi            2    chewy  soft

kacang         2    no nuts  nuts


                           Number of Observations Read             8
                           Number of Observations Used             8

Identity(rating)
           Algorithm converged.


          The TRANSREG Procedure Hypothesis Tests for Identity(rating)


                        Root MSE           0.50000    R-Square    0.9500
                        Dependent Mean     7.00000    Adj R-Sq    0.9125
                        Coeff Var          7.14286


                              The TRANSREG Procedure

             Utilities Table Based on the Usual Degrees of Freedom
Importance
Standard    (% Utility Label   Utility       Error        Range)    Variable
Intercept           7.0000     0.17678                  Intercept

coklat, dark       -1.2500     0.17678        50.000    Class.coklatdark
coklat, milk        1.2500     0.17678                  Class.coklatmilk

isi, chewy          0.5000     0.17678        20.000    Class.isichewy
isi, soft          -0.5000     0.17678                  Class.isisoft

kacang, no nuts    -0.7500     0.17678        30.000    Class.kacangno_nuts
kacang, nuts        0.7500     0.17678                  Class.kacangnuts
```

Nampak dari output pernyataan Algorithm converged artinya tidak ada masalah dalam iterasi, nampak pula nilai R^2=0.95. pada tabel terakhir

tabel utilities menampilkan yang nilai kegunaan dari jenis permen, mulai dari yang paling banyak disukai hingga yang paling sedikit disukai.Nilai positif dalam kolom utility lebih disukai dibanding nilai negatifnya.Coklat milk lebih disukai dibanding coklat dark, isi chewy lebih disukai dibanding soft, kacang lebih disukai dibanding tanpa kacang.

CA menyediakan sebuah perkiraan dekomposisi dari nilai rating awal. Perkiraaan nilai kegunaan (utility) dari sebuah permen adalah jumlah intercept dan part-worth utility. CA model untuk preferensi coklat tipe ke-i, isi ke-j, dan kacang ke-k. adalah

$$y_{ijk} = \mu + \beta_{1i} + \beta_{2j} + \beta_{3k} + \epsilon_{ijk}$$

Sehingga untuk i =1,2; j=1,2; k=1,2; dengan

$$\beta_{11} + \beta_{12} = \beta_{21} + \beta_{22} = \beta_{31} + \beta_{32} = 0$$

Part-worth utilities untuk level dari atribut adalah nilai harapan parameter $\hat{\beta}_{11}, \hat{\beta}_{12}, \hat{\beta}_{21}, \hat{\beta}_{22}, \hat{\beta}_{31}, \hat{\beta}_{32}$; adapun nilai harapan dari intercept adalah μ dan komponen galat adalah ϵ_{ijk}

Nilai dugaan utility dari kombinasi ijk adalah

$$\hat{y}_{ijk} = \hat{\mu} + \hat{\beta}_{1i} + \hat{\beta}_{2j} + \hat{\beta}_{3k}$$

Sebagai contoh kombinasi yang paling disukai milk chewy nuts dapat diduga dengan nilai utility-nya.

$$7{,}0+1{,}25+0{,}5+0{,}75=9{,}5= \hat{y} \approx y = 9{,}0$$

Dan yang kombinasi yang paling tidak dipilih adalah kombinasi dark soft no nuts, nilai dugaan utility dan nilai sebenarnya adalah sebagai berikut

$$7{,}0 - 1{,}25 - 0{,}5 - 0{,}75 = 4{,}5 = \hat{y} \approx y = 4{,}0$$

Nilai dugaan utilities adalah nilai dugaaan regresi; kuadrat korelasi antara dugaan utilities untuk setiap kombinasi dan nilai rating preference sebenarnya adalah R^2.

Kolom important dihitung dari jarak antar faktor (atribut) di kolom utility (part worth utility).Setiap jarak dikalikan 100 lalu dibagi jumlah dari semua jarak.Faktor yang memiliki jarak tertinggi adalah yang paling penting dalam menentukan preferensi. Namun ketika faktor memiliki level faktor yang beragam kadang kala dapan menaikkan nilai importence (lihat wittink, krishnamurthi, dan Reibstein, 1989).

Dari nilai importance menunjukkan nampak bahwa type coklat, dengan nilai importance 50% adalah atribut yang paling penting dalam penentuan preferensi.

$$\frac{100 \times (1.25 — 1.25)}{(1.25 — 1.25) + (0.5 — 0.5) + (0.75 — 0.75)} = 50\%$$

Nilai kedua yang paling penting adalah permen yang berisi kacang, dengan nilai nilai importance 30%

$$\frac{100 \times (0.75 — 0.75)}{(1.25 — 1.25) + (0.5 — 0.5) + (0.75 — 0.75)} = 30\%$$

Dan isi permen dengan nilai importance 20% adalah yang paling kecil nilai kepentingannya

$$\frac{100 \times (0.50 — 0.50)}{(1.25 — 1.25) + (0.5 — 0.5) + (0.75 — 0.75)} = 20\%$$

17.3.2. Non-Metric Conjoint Analysis

Dalam contoh ini,PROC TRANSREG digunakan untuk melakukan analisis conjoint nonmetrik dari dataset permen coklat.perbedaannya terletak pada tranformasi variabel respon. dalam analisis conjoint nonmetrik menggunakan tranformasi monotone terhadap variabel rating. Juga kita tidak menggunakan option short sehingga kita dapat melihat tabel sejarah iterasi. Adapun pernyataan output digunakan agar nilai rating yang telah transformasikan tersimpan diluar output data set.

```
odsexclude notes anova liberalanova conservanova
mvanova liberalmvanova conservmvanova;
proctransregutilitiesseparators=', ';
        title2'NonMetric Conjoint Analysis';
        model monotone(rating)=class(coklat isi kacang / zero=sum);
output;
run;
```

Analisis conjoint nonmetrik secara iteratif menurunkan tranformasi monotone dari variabel rating .

The TRANSREG Procedure

Dependent Variable Monotone(rating)

Class Level Information

Class	Levels	Values
coklat	2	dark milk
isi	2	chewy soft
kacang	2	no nuts nuts

Number of Observations Read	8
Number of Observations Used	8

TRANSREG Univariate Algorithm Iteration History for Monotone(rating)

Iteration Number	Average Change	Maximum Change	R-Square	Criterion Change	Note
1	0.08995	0.23179	0.95000		
2	0.01263	0.03113	0.96939	0.01939	
3	0.00345	0.00955	0.96981	0.00042	
4	0.00123	0.00423	0.96984	0.00003	
5	0.00050	0.00182	0.96985	0.00000	
6	0.00021	0.00078	0.96985	0.00000	
7	0.00009	0.00033	0.96985	0.00000	
8	0.00004	0.00014	0.96985	0.00000	
9	0.00002	0.00006	0.96985	0.00000	
10	0.00001	0.00003	0.96985	0.00000	Converged

Algorithm converged.

The TRANSREG Procedure

The TRANSREG Procedure Hypothesis Tests for Monotone(rating)

Root MSE	0.38829	R-Square	0.9698
Dependent Mean	7.00000	Adj R-Sq	0.9472
Coeff Var	5.54699		

Utilities Table Based on the Usual Degrees of Freedom

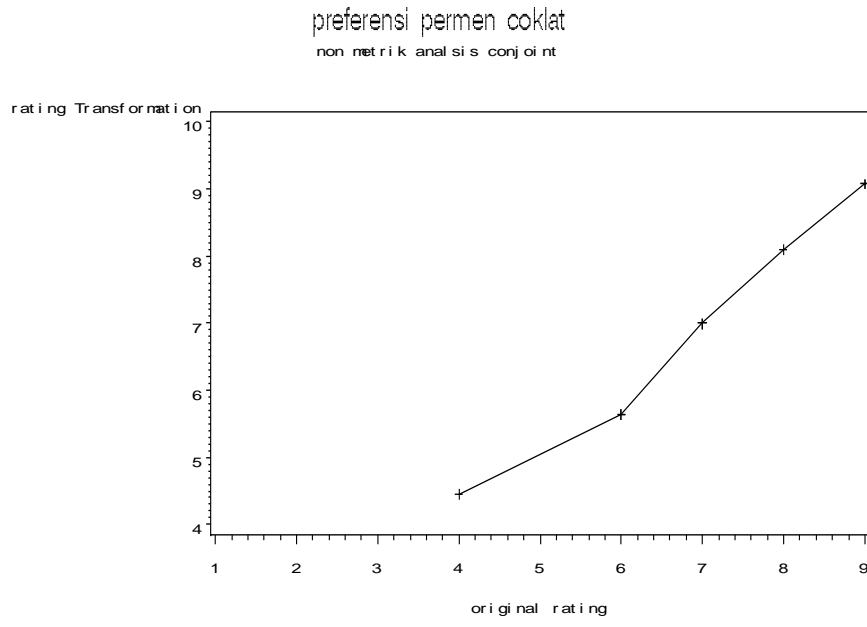
Label	Utility	Standard Error	Importance (% Utility Range)	Variable
Intercept	7.0000	0.13728		Intercept
coklat, dark	-1.3143	0.13728	53.209	Class.coklatdark
coklat, milk	1.3143	0.13728		Class.coklatmilk
isi, chewy	0.4564	0.13728	18.479	Class.isichewy
isi, soft	-0.4564	0.13728		Class.isisoft
kacang, no nuts	-0.6993	0.13728	28.312	Class.kacangno_nuts
kacang, nuts	0.6993	0.13728		Class.kacangnuts

The standard errors are not adjusted for the fact that the dependent variable was transformed and so are generally liberal (too small).

Nilai R^2 naik dari 0.95 menjadi 0,969585 dalam kasus non metrik ini.nilai importance dan utility (part-worth utility hanya sedikit perbedaannya dibanding analisis metrik. Namun secara umum pola dari hasilnya relatif sama.

Prosedur GPLOT selanjutnya digunakan untuk memplot tranformasi dari variabel rating.

```
procsort; by rating; run;
procgplot;
        title h=1.5'preferensi permen coklat';
        title2 h=1'non metrik analsis conjoint';
        plot trating * rating = 1 / framehaxis=axis2 vaxis=axis1;
        symbol1v=plus i=join;
        axis1order=(1 to 10)
                label =(angle=90'tranformation variabel rating';
        axis2order=(1 to 9) label=('original rating');
        run; quit;
```

dan diperoleh output:

Dalam hal ini.transformasihampir mendekati linear. Dalam kasus lain, nilai R^2 dapat naik lebih dari seperti dalam teladan ini. dan transformasi mungkin nampak dengan jelas nonlinear.

DAFTAR PUSTAKA

Borg Ingwer, Groenen J. F. Patrick. 2005. *Modern Multidimensional Scaling:Teori and Applications*.Berlin Springer.

Curry, J. 1996. *Understanding Conjoint Analysis in 15 Minutes*, Sawtooth Software

Cox F. Trevor, Cox A.A Michael. 2000. *Multidimensional Scaling.* Washington DC Chapman dan Hall CRC.

Gabriel, R. 1971. The Biplot Graphic Display of Matrices with Application to Principal Component Analysis.Journal of Biometrika, 58,3: 453-467.

Gittins,R.1984, Canonical Analysis A Review with Applications in Ecology, Springer-Verlag

Gunarto, Muji. 2008. *Membangun Model Persamaan Struktural (SEM) dengan LISREL 8.30*. [online]

Hair, Joseph F et. all. 2006. *Multivariate Data Analysis*. Sixth Edition. New Jersey: Pearson Prentice Hall

Jollife IT. 2002. Primcipal Component Analysis. Second Edition. Springer-Verlag, New York.

Johnson, R A. and Wichern,D W., 2007. *Applied Multivariate Statistical Analysis 6th Edition*, Prentice Hall

Junaidi. 2008. *SEM dan LISREL (Seri 1-8)*. [Online]

Kuhfeld, W F, Conjoint analysis, article. http://support.sas.com/techsup/ tnote/ tnone_stat.html#market

Lebart, L,Morineau, A, Warwicck, K.M. 1984. *Multivariate Descriptive Statistically Analysis*. John Wiley & Sons, New York.

Nugroho, S, 2008. *Statistika Multivariate Terapan*, UNIB Press

Orme, B. 2004. *Conjoint Analysis: Thirty-Something and Counting,* Sawtooth Software, Inc

Rencher, A C.,2002. *Methods of Multivariate Analysis 2nd Edition*, Wiley Interscience

Wijanto, Setyo Hari. 2008. *Structural Equation Modeling dengan LISREL 8.8: Konsep dan Tutorial*. Yogyakarta: Graha Ilmu